Role of Mimivirus Collagen in Rheumatoid Arthritis

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Nikunj Shah

aus

Indien

Promotionskomitee

Prof. Dr. Thierry Hennet (Vorsitz)

PD. Dr. Lubor Borsig

Prof. Dr. Matthias Baumgartner

PD. Dr. Patricie Burda

Zürich, 2013

Table of Contents

Summary

Giant viruses with genomes larger than several bacterial genomes have been extensively described in the past decade. These giant viruses express proteins previously not identified in any viruses and thus giant viruses have garnered widespread evolutionary and biological importance. The giant mimivirus is an amoeba-associated virus that replicates in the ubiquitously present amoeba *Acanthamoeba polyphaga*. Sequence analysis revealed that mimivirus harbours a linear double stranded DNA genome of ~1.2 Mbp making mimivirus the third largest viral genome known to mankind. Mimivirus exhibits many features that have never been identified in a virus before, such as genes for aminoacyl-tRNA synthetases, DNA repair mechanisms, and enzymes related to novel metabolic pathways. Interestingly, mimivirus also harbours seven open reading frames encoding collagen-like proteins sharing structural similarity with collagen proteins in animals. Infections with microbes having similar immunogenic epitopes as self-proteins have been linked with the pathogenesis of autoimmune diseases by a phenomenon known as molecular/antigen mimicry. Considering the broad occurrence of giant viruses like mimivirus in various ecosystems, humans are at a potential risk of exposure to them. Here, we investigated whether exposure of humans to mimivirus leads to the production of antibodies towards viral collagen which might cross-react with endogenous collagen leading to molecular/antigen mimicry thereby establishing autoimmunity.

Firstly, using protein biotinylation we showed that mimivirus collagen protein L71 is found at the surface of mimivirus particles among 60 other proteins. To investigate the potential of mimivirus collagens to induce an autoimmune response, we used a mouse model of collagen induced arthritis, which closely resembles the autoimmune disease of rheumatoid arthritis in humans. When DBA/1 mice were immunized intra-dermally with mimivirus proteins, production of anti-collagen antibodies targeting mouse collagen type II was elicited as measured by ELISA in blood sera. This antibody response was accompanied by T cell reactivity to collagen as analysed in recall assays. Cells isolated from draining lymph nodes of immunized mice were shown to proliferate when stimulated with mouse, bovine and mimivirus L71 collagen fragments demonstrating presence of auto-reactive T cells against collagen. Joint inflammation and cartilage

destruction were also observed by clinical scoring, histology and *in vivo* fluorescence imaging techniques. Besides, we demonstrate IgG and IgM titers against mimivirus proteins in a pool of blood sera from human healthy subjects and rheumatoid arthritis patients suggesting human exposure and resulting in immune response mounted towards mimivirus. Mass spectrometric analyses after immunoprecipitation of mimivirus proteins to IgG from human sera, confirmed that several mimivirus proteins are recognized by human sera. These viral proteins include among others the major capsid protein (L425), core protein (L410) and GMC type oxidoreductase (R135). Interestingly, the most frequent mimivirus proteins recognized by human sera were surface proteins according to our surface biotinylation study. Western blot analyses with purified His_6-tagged recombinant mimivirus capsid protein L425 and collagen protein L71 further validated the antibody response to specific mimivirus proteins. We observed that, 30% of healthy subjects and 36% of rheumatoid arthritis sera recognized the major mimivirus capsid protein L425. By contrast, only 6% of healthy subject sera recognized the mimivirus collagen protein L71, whereas 22% of rheumatoid arthritis sera were positive for L71.

The work presented in this thesis, shows that mimivirus collagen can trigger the development of arthritis in a mouse model, thereby confirming the ability of mimivirus collagen to induce T and B cell responses cross-reacting with self-epitopes. A plausible association between exposure to mimivirus and autoimmunity was corroborated by the frequent occurrence of antibodies against mimivirus collagen in rheumatoid arthritis patients. Repeated exposure to mimivirus leads to antibody formation to virus collagen thereby breaking immune tolerance for endogenous collagens. These findings cast a new light on the environmental exposure to giant viruses representing a potential risk factor in triggering autoimmunity to collagens in humans.

Zusammenfassung

So genannte Riesenviren, deren Genomgrösse jene von vielen Bakterien übertreffen, wurden in der letzten Dekade ausführlich untersucht. Diese Riesenviren exprimieren Proteine, die man bisher in Viren nicht beobachten konnte, und die somit hinsichtlich Evolution und Biologie von weitreichender Bedeutung sind. Das zu den Riesenviren gehörende Mimivirus repliziert sich mit Hilfe der Amöbe *Acanthamoeba polyphaga*. Mittels Sequenzanalysen der Mimivirus-DNS wurde eine lineare, doppelsträngige DNS mit ~1.2 Mbp gefunden, die das bisher drittgrösste virale Genom darstellt. Von Viren unbekannte Gene wurden identifiziert, darunter Gene für Aminoacyl-tRNA-Synthetasen, für die DNS Reparaturmaschinerie, und für Enzyme von Stoffwechselwegen. Das Mimivirus beherbergt auch sieben offene Leserahmen, die Kollagen-ähnliche Proteine kodieren, die eine strukturelle Ähnlichkeit mit tierischen Kollagenen aufweisen. Die molekulare Mimikry beschreibt die Angleichung der Proteine oder Kohlenhydrate eines Pathogens an jene des Wirts, um bei der Infektion einer Immunreaktion des Wirts zu entgehen. Oft führt dies zu einer Autoimmunantwort des Wirts auf eigene Epitope. Da Riesenviren wie das Mimivirus in verschiedenen Ökosystemen auftreten, sind Menschen einem potentiellen Risiko eines Kontakts ausgesetzt. In dieser Doktorarbeit untersuchen wir, ob Kontakt zwischen Mensch und Mimivirus zu einer Produktion von Antikörpern gegen das virale Kollagen führt und ob ein solcher Kontakt aufgrund molekularer Mimikry eine mögliche Antikörper-Kreuzreaktion mit endogenem Kollagen nach sich zieht, was zur Entstehung einer Autoimmunität führen kann.

Mittels Biotinylierung von Oberflächenproteinen konnten wir Mimivirus-Kollagen L71 unter 60 anderen Proteinen auf der Hülle des Mimivirus nachweisen. Zur Untersuchung des Potenzials des Mimivirus-Kollagens, eine Autoimmunreaktion auszulösen, wurde ein Mausmodell verwendet, in dem Arthritis mittels Kollageninjektion ausgelöst wird, die der menschlichen Rheumatiden Arthritis gleicht. Die verwendeten DBA/1 Mäuse wurden intra-dermal mit Mimivirus-Proteinen immunisiert und die Blutseren auf eine hervorgerufene Produktion von anti-Kollagen Antikörpern gegen das Maus-Kollagen Typ II wurden mittels ELISA getestet. Die Antikörperreaktion wurde von einer T-Zellreaktivität begleitet, die wir *in vitro* mit isolierten T-Zellen nachweisen konnten. Eine Stimulation von isolierten Immunzellen aus Lymphknoten von immunisierten

Mäusen mit Mauskollagen, bovinem Kollagen und Fragmenten des Mimivirus-Kollagen L71 löste eine Zellproliferation aus, womit die Präsenz von auto-reaktiven T-Zellen gegen Kollagen bewiesen werden konnte. Die Entzündung der Gelenke und die Knorpelzerstörung wurden mit einer klinischen Skala, Histologie und mit einer Fluoreszenz-basierten bildgebenden Methode *in vivo* quantifiziert. Darüber hinaus konnten wir in Blutseren der gesunden Kontrollgruppe und der Rheumatiden Arthritis-Gruppe IgG und IgM Titer gegen Mimivirusproteine detektieren, was humanen Kontakt mit dem Mimivirus und eine nachfolgende Immunantwort nachweist. Massenspektrometrische Analysen von Immunopräzipitationen von Mimivirus-Proteinen mit humanem IgG aus Blutserum bestätigten die Antikörperproduktion gegen verschiedene Mimivirus-Proteine. Unter anderen detektierten Proteinen befinden sich das Hauptkapsid-Protein L425, das Kernprotein L410 und eine Oxidoreduktase vom Typ GMC, R135. Interessant dabei war, dass sich die von den Seren am häufigsten erkannten viralen Proteine auf der Virusoberfläche befinden, wie die Biotinylierungs-Experimente der Oberflächenproteine bestätigten. Westernblotanalysen mit rekombinanten, mit einem His_6-tag versehenen L425 und Kollagen L71 validierten die spezifische Immunantwort auf Mimivirus-Proteine. Das Kapsidprotein L425 wurde von 30% der Blutseren der gesunden Kontrollgruppe und von 36% der Blutseren der Rheumatiden Arthritis-Gruppe erkannt. Nur 6% der Blutseren der gesunden Kontrollgruppe hingegen erkannten das Kollagen L71, während 22% der Blutseren der Rheumatiden Arthritis-Gruppe L71 erkennen konnten.

In dieser Dissertation wird dargelegt, dass Mimivirus-Kollagen in einem Mausmodell Arthritis auslösen kann und dieser Krankheitsverlauf von einer Induktion der T- und B-Zellantwort begleitet wird, die eine resultierende Kreuzreaktion mit endogenem Kollagen ermöglicht. Die vermehrte Präsenz von Antikörpern gegen Mimivirus-Kollagen in Rheumatiden Arthritis-Patienten erhärtet den Zusammenhang zwischen Kontakt mit Mimivirus und Autoimmunreaktionen. Wiederholter Kontakt mit Mimiviren löst demnach eine Antikörperproduktion gegen virales Kollagen aus, was zu einer Immunintoleranz gegenüber endogenem Kollagen führt. Diese Resultate beleuchten einen neuen Aspekt des umgebungsbedingten Kontakts von Menschen mit Mimiviren, was einen potentiellen Risikofaktor für die Auslösung einer Autoimmunität gegen endogenes Kollagen darstellt.

Abbreviations

APC	Antigen presenting cell
bp	Base pair
CD	Cluster of differentiation
CIA	Collagen induced arthritis
CII	Collagen type II
CroV	*Cafeteria roenbergensis* virus
dsDNA	Double stranded DNA
ER	Endoplasmic reticulum
Ig	Immunoglobulin
IL	Interleukin
kbp	Kilo base pair
MHC	Major histocompatibility complex
MMPs	Matrix metalloproteinases
NCLDV	Nucleo-cytoplasmic large DNA viruses
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ORF	Open reading frame
PBCV-1	*Paramecium bursaria* chlorella virus 1
PCR	Polymerase chain reaction
RANKL	Receptor activator of NF-κB ligand
TCR	T-cell receptor
TNF	Tumour necrosis factor

Introduction

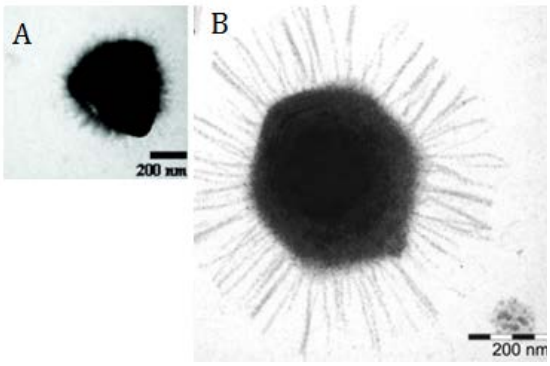
Infections with foreign agents having similar immunogenic epitopes as auto-antigens have long been postulated and linked with the pathogenesis of many autoimmune diseases by a phenomenon known as molecular/antigen mimicry. Giant viruses like mimivirus have been recently isolated from natural environments and garnering extensive evolutionary and biological importance due to the presence of unique genes encoding for proteins which have never been identified in a virus before. Many of the mimivirus proteins such as collagen-like proteins are similar to as in many other eukaryotes. Here, we postulated the possibility of mimivirus possessing collagen-like proteins to potentially trigger an autoimmune disease like rheumatoid arthritis.

Acanthamoeba polyphaga mimivirus

Discovery

The putative cause of a hospital pneumonia outbreak in 1992, in Bradford, England was identified to be a microorganism growing in amoebae and resembled a Gram-positive coccus and was named *Bradfordcoccus* [1]. However, after several attempts of different extraction protocols and low-stringent polymerase chain reactions, no bacterial DNA was isolated using the universal 16S ribosomal DNA bacterial primers [1]. It was much later in 2003, by electron microscopy revealed that this microorganism showed a typical viral morphology and mature particles with ~400 nm in diameter, icosahedral capsid symmetry without envelope but with the presence of 80-100 nm fibrils attached to its capsid (Fig. 1) [1]. Since this viral particle resembled a bacterium on Gram-staining it was named Mimivirus- for mimicking microbe. It was also identified that the natural host of this virus is the ubiquitously present unicellular amoeba *Acanthamoeba polyphaga* and hence the complete scientific name of the virus *Acanthamoeba polyphaga* mimivirus or simply mimivirus.

Fig. 1-Electron microscopy of mimivirus. (Figure adapted from [1,2]) (**A**) Initial electron microscopy revealing the icosahedral symmetry of mimivirus. (**B**) Transmission electron microscopy of isolated mimivirus particle showing the icosahedral symmetry of the capsid (~400 nm) surrounded by ~80-100 nm fibres around it harbouring the dense genetic core within it.



Classification

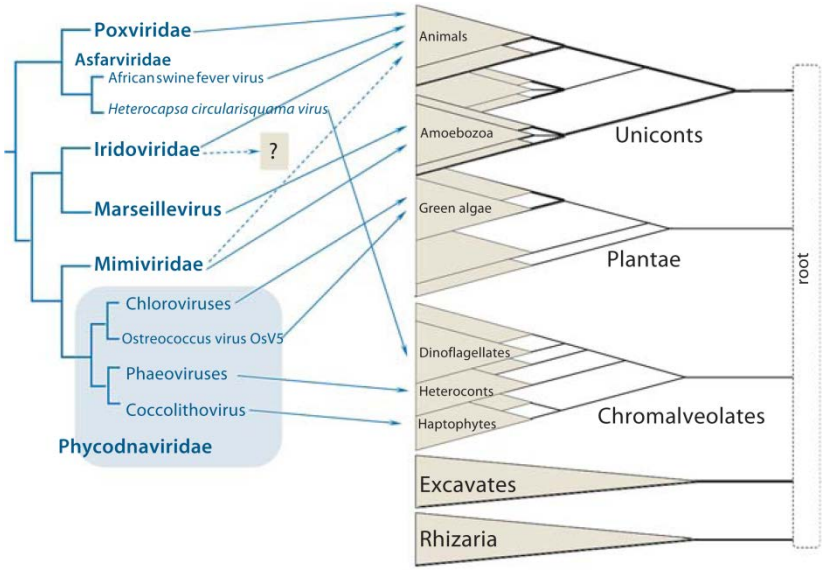
According to the International Committee on Taxonomy of Viruses, so far, mimivirus has not been assigned to any phylum, class or order but has been placed in a newly formed family called *Mimiviridae*. Furthermore, after whole sequence analysis of mimivirus revealing its ~1.2 Megabasepair (Mbp) linear DNA genome [3], it has been placed in the monophyletic nucleo-cytoplasmic large DNA virus (NCLDV) group which includes six other families of large DNA viruses which share many of the structural and genomic characteristics [4,5].

Nucleo-Cytoplasmic Large DNA Virus (NCLDV)

Viruses are generally known to be obligate intracellular parasites and depend on the hosts for protein synthesis and might also depend entirely or in part for transcription and replication [6]. All viruses share two common properties, namely, obligatory intracellular parasitism and encapsidation of its either RNA or DNA genome into virus particles known as virions [6]. Other than these common properties, viruses show great diversity in every regard including host interactions, genome size, virion structure, replication cycles and capsid structure [6]. Viruses are known to infect and propagate in all cell types from all the three domains of life (archaea, bacteria and eukarya). Using modern comparative genomic tools, seven families of large and giant DNA viruses infecting various eukaryotes have been placed in a common monophyletic group of viruses denoted as NCLDVs [4,5]. These seven families are- *Poxviridae*, *Asfaviridae*,

Iridoviridae, *Phycodnaviridae*, *Ascoviridae*, *Mimiviridae* and putative new family *Marseilleviridae*. By comparative genomic studies, it appears that these seven families infecting wide range of eukaryotes to have a common ancestry (Fig. 2) [4,5].

A



B

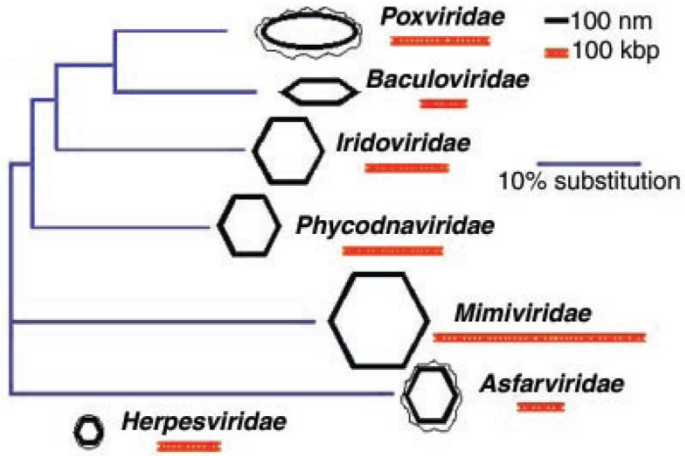


Fig. 2-Phylogenetic trees of the NCLDV. (Figure adapted from [1,7]) (**A**) Cross-mapping of phylogenetic trees of NCLDV and eukaryotes showing established virus-host relationships (solid lines) and putative virus-host relationships (broken lines) as analysed by metagenomic studies. The broken line from *Mimiviridae* clearly shows a possibility of mimivirus establishing a relationship with animals and the broken line from *Iridoviridae* shows possibility of them infecting marine eukaryotic hosts but is still unclear. (**B**) Mimivirus appearing as a deep branch in the phylogenetic tree from alignment studies of ribonucleotide reductase small subunit indicating early divergence from other virus families.

This main group of viruses have been designated as NCLDVs since they all have double-stranded DNA (dsDNA) genomes (either linear or circular) and they all either replicate completely in the cytoplasm of the host or start their replication cycle in the nucleus of the host and end in the cytoplasm [4,5]. These viruses have a broad range of hosts, ranging from protists to algae to plants to even animals (Fig. 2 and Table 1). Moreover, they either have shared or exclusive features of virion structures and genomic DNA [5]. There seems to be a common ancestry of these seven families and is shown by the

presence of certain set of genes that are conserved in all or nearly all members of this NCLDV group [7,8]. These conserved set of genes encode proteins that helps the virus to propagate in the hosts and includes genes for replication and recombination, virion structure and assembly and transcription and nucleic acid processing [7].

Table 1-Seven families of NCLDV and some properties (table adapted from [7])

Family	**Genome range (kbp)**	**Genome structure**	**Replication site**	**Hosts**
Poxviridae	130-380	Linear	Cytoplasm	Insects, reptiles, birds, mammals
Asfaviridae	170	Linear	Cytoplasm	Mammals
Ascoviridae	150-190	Circular	Nucleus/cytoplasm	Insects
Iridoviridae	100-220	Linear	Nucleus/cytoplasm	Insects, vertebrates
Phycodnaviridae	150-400	Linear/circular	Nucleus/cytoplasm	Green algae
Mimiviridae	700-1200	Linear	Cytoplasm	Amoebae
Marseilleviridae	370	Circular	Nucleus/cytoplasm	Amoebae

Poxviridae

This family of viruses are subdivided into two subfamilies, *Chordopoxvirinae* which are known to infect vertebrates (reptiles, birds and mammals) and *Entomopoxvirinae* which generally propagate in invertebrates (insects) [9,10]. The most common poxvirus is the human pathogen Variola virus which causes smallpox and also the Vaccinia virus which was used as a biotechnological tool in virology to produce a vaccine against smallpox which is now successfully eradicated [9,10]. Poxviruses have a linear dsDNA genome

and complex unusual ovoid or brick-shaped enveloped virions (~140-260 nm in diameter).

Asfaviridae

The typical member of this group of viruses is the African swine fever virus (ASFV) and is known to be the causative agent of fatal haemorrhagic disease in domestic swine [11]. Like the poxviruses, they too have a linear dsDNA genome, but an icosahedral nucleocapsid and an envelope (~80 nm in diameter). Recently, a marine virus *Heterocapsa circularisquama*, was identified which infects a dinoflagellate and partial genome sequencing indicated similarity to ASFV [12]. This shows that the initial rational that asfaviruses only infects terrestrial animals does not hold true and that this family of viruses seems to have a broad range of eukaryotic hosts [12].

Ascoviridae

Ascoviruses have an intra-capsid lipid membrane and an envelope and are similar to poxviruses in virion structures in being ovoid (~130 nm in diameter) and infect insects [13]. It seems that, due to the ovoid virions, ascoviruses and poxviruses most likely are a case of evolutionary convergence [7,8]. The type member of this group is *Spodoptera fruigiperda ascovirus* 1a and infects insects and is known to cause chronic fatal disease which manipulates apoptosis for its propagation [14].

Iridoviridae

Iridoviruses have an icosahedral nucleocapsid and an internal lipid membrane (~120-150 nm in diameter) and mainly infects insects [15]. Some iridoviruses are known to infect non-mammalian vertebrates (fish, amphibians and reptiles) and these viruses generally have an external envelope around its capsid. Apparently, ascoviruses have evolved from iridoviruses [8]. Common examples are ranavirus and megalocytivirus which are contributing in declining the amphibian population and having adverse effects on acquaculture [15].

Phycodnaviridae

Phycodnaviruses have icosahedral nucleocapsid virions (~130-200 nm in diameter), similar to asfaviruses, however they do not have any envelope [16]. They mainly infect green algae and play a role in the ecology of freshwater and marine environments [16]. The widely studied virus of this group is *Paramecium bursaria* chlorella virus 1 (PBCV-1) which infects the symbiotic chlorella algae in freshwater environments. PBCV-1 showed features which were unique in viruses such as presence of DNA-endonucleases, intron-splicing genes and many enzymes involved in glycosylation machinery which were never identified in a virus before and thus showing a long evolutionary history [16,17]. Much pioneering work regarding the first virally encoded protein glycosylation machinery has been accomplished.

Glycosylation machinery of PBCV-1:

Many structural proteins in viruses such as herpesviruses, poxviruses, paramyxoviruses and in PBCV-1 are glycosylated i.e. post-translationally modified by the addition of sugar groups (known as glycans) by designated enzymes termed as glycosyltransferases [18,19]. Classically host-encoded glycosyltransferases located in the endoplasmic reticulum (ER) and Golgi apparatus have been known to glycosylate viral proteins [18,19]. However, in PBCV-1 seven putative glycosyltransferases were identified in its genome and this was the first case of virally encoded putative glycosyltransferase which help the virus in glycosylating its own major capsid protein (Vp54) [20]. Interestingly, it was identified that none of the seven putative glycosyltransferase-encoding genes have a signal peptide which would target them to the ER [20]. Furthermore, it has been known that compounds, even at doses lethal to host algae, such as tunicamycin which inhibits ER-Golgi localized N-linked protein glycosylation and Brefeldin A which inhibits protein transport from ER to Golgi does not affect PBCV-1 replication [21,22]. Seven neutral sugars- glucose, fucose, galactose, mannose, xylose, rhamnose and arabinose have been identified in the glycan structures of Vp54 [20,23] and it was predicted that one of these seven glycosyltransferases is a mannosyltransferase [20]. A common sugar N-acetylglucosamine is commonly found in many N and O-linked glycan chains but was however absent in Vp54 glycans [20,23]. The cytoplasmic location of these glycosyltransferases suggested that Vp54 is glycosylated in the cytoplasm and thus is a process different from ER-Golgi glycosylation [20]. In addition, these

glycosyltransferases show sequence homology to bacterial enzymes, further confirming the ancestral origin existing before ER-Golgi formation [20]. Thus the glycosylation of its own capsid protein is independent of the host algal ER-Golgi machinery.

Moreover, PBCV-1 also encodes for various other enzymes involved in sugar metabolism such as hyaluronan synthase and hyaluronan precursors [20,23]. Until the discovery of these genes in PBCV-1, it was known that the linear polysaccharide hyaluronan was occurring only in vertebrates as a constituent of extracellular matrix and certain pathogenic bacteria which forms a protective hyaluronan capsule to evade the host immune system [24]. Upon infection with PBCV-1, host cells were covered with hyaluronan synthesized by these enzymes which is a survival tactic for the virus since the polysaccharide layer prevented the uptake of virus-infected algae by the protist paramecium [17,20].

Mimiviridae

The current record holders in the viral kingdom for both particle and genome size, which exceeds various bacteria and simple prokaryotic organisms belong to the *Mimiviridae* family [3,7]. Genomes of more than 1 Mbp encoding some genes never seen in a virus before as in mimivirus [3], mamavirus [25], megavirus [26] which infect amoebae have brought to light the question of evolution of viruses and whether viruses should be included in the tree of life [27]. Members of NCLDV were known as large viruses, but after the discovery of mimivirus revealing its genome sequence of ~1.2 Mbp and particle size of ~700 nm in diameter, these viruses came to be known as giant viruses (or girus) [28]. Mimivirus has an icosahedral capsid (400-500 nm in diameter) and additional filamentous fibres (~80-100 nm) attached to its surface [1,3]. Many mimivirus-like sequences have been identified by marine metagenomics suggesting that viruses of this family infect a broad range of planktonic unicellular eukaryotes [29].

Marseilleviridae

Marseillevirus and Lausannevirus, two new giant viruses infecting amoebae as well were recently identified and a putative novel family *Marseilleviridae* is postulated to best fit these viruses [30-32]. Marseillevirus also has an icosahedral capsid (~250 nm in diameter) surrounded by fibres but in contrast to most NCLDVs possesses a circular

DNA genome and thus has been suggested to create a novel family [30]. Sequence analysis of Marseillevirus revealed a ~368 kbp genome, with much closer similarity to mimivirus genome and certain genes common as in iridoviruses and ascoviruses [30]. Marseillevirus also encodes some proteins which are previously not yet identified in other NCLDVs such as three histone-like proteins potentially facilitating condensing and packaging of viral DNA [30].

Genome complexity of mimivirus

The capsid of mimivirus harbours a linear double stranded DNA genome of 1,181,404 bp which is larger than several cellular genomes (GenBank Accession number-NC_014649) [3]. Considering the large genome of approximately 1.2 Mbp and a characteristic large capsid diameter of at least 400-500 nm, mimivirus has been termed as a giant virus (or girus) [28]. The genome encodes for 1262 putative open reading frames (ORFs) of more than 100 amino acid residues and out of which at least 979 are expected to be protein-coding genes [3]. Mimivirus genome harbours some genes of all the four classes which are common to all the NCLDVs however there are a few genes unique to mimivirus and has never been identified in a virus ever before. A part of transcription of mimivirus proteins occurs in the cytoplasm of the host since the same transcription-related core genes without the exception of RNA polymerase subunit 10 is present as found in *Poxviridae* and *Asfaviridae* [7,8]. The genome sequencing of mimivirus revealed some unique features among dsDNA viruses and certain new genes (Table 2). Firstly, a virus is known to depend on its hosts for protein translation and synthesis and is the major characteristic that differentiates virus from cellular 'living' organisms. However, there have been some viruses wherein transfer-RNA (tRNA)-like genes have been identified such as in bacteriophage T4 [33], herpes virus [34], and PBCV-1 [35]. Similarly, mimivirus also possesses six tRNA-like genes. Moreover, surprisingly the genome of mimivirus encodes for some genes which have never been identified in a virus and is known to be a trademark of cellular organisms such as at least four aminoacyl-tRNA synthetases, certain translation initiation and elongation factors and peptide chain release factors [3].

Table 2-Major new genes identified in mimivirus genome (table adapted from [3])

Mimivirus ORF	**Putative function**	**Remarks**
R663	Arginyl-tRNA synthetase	Translation
L124	Tyrosyl- tRNA synthetase	Translation
L164	Cysteinyl- tRNA synthetase	Translation
R639	Methyonyl- tRNA synthetase	Translation
R726	Peptide chain release factor	Translation
R464, L496	Translation initiation factor	Translation
L359	DNA mismatch repair	DNA repair
R693	Methylated-DNA-protein-cysteine methyltransferase	DNA repair
R406	Alkylated DNA repair	DNA repair
L687	Endonuclease for UV-irradiated DNA repair	DNA repair
R194, R480, L221	Topoisomerase I and II	DNA accessibility
R418	Synthesis of nucleoside triphosphates	Metabolism
R565, L716	Glutamine synthesis	Metabolism
R689	N-acetylglucosamine-1-phosphate, uridyltransferase	Polysaccharide synthesis
L136	Sugar transaminase, Viosamine biosynthesis	Polysaccharide synthesis
L612	Mannose-6-phosphate isomerase	Glycosylation
R141	Sugar nucleotide (L-Rhamnose) biosynthesis	Glycosylation
L230	Pro-collagen hydroxylase and glycosyltransferase enzyme	Glycosylation

Secondly, mimivirus genome is well prepared to repair some DNA mismatches caused by various factors such as ultraviolet light, ionizing radiations, chemical mutagens, methylating and alkylating agents. For example, genes encoding glycosylases and endonucleases have been identified in the genome to locate and repair oxidized purines and ultraviolet induced DNA damages [3]. However, even though repair genes and enzymes are present, mimivirus can be quickly killed by irradiation with gamma rays or 15 min exposure to ultraviolet light [3].

Thirdly, enzymes responsible for fixing problems associated with replication, recombination and transcription such as DNA topoisomerases [36] are also found in the mimivirus genome. Mimivirus and several other NCLDV members except poxviruses encode their own type IIA topoisomerases [3]. Surprisingly, mimivirus not only encodes type IIA and type IB topoisomerases but also shows type IA topoisomerases which have never been reported in a virus [3,36]. Furthermore, many genes revealed that mimivirus has its own host-independent glycosylation machinery, like PBCV-1, helping the virus glycosylating its many structural proteins including its major capsid protein L425 (explained in detail later).

Overall the mimivirus genome shows certain features common with other dsDNA viruses in the NCLDV group however surprisingly the presence of mainly the protein translation machinery also shows the divergence from classically defined host protein synthesis machinery dependent viruses.

Structure of mimivirus particles

The early electron micrograph of the pneumonia causing microorganism showed a typical viral icosahedral morphology confirming the infectious agent to be a virus (Fig. 1A) [1]. Further advanced microscopic techniques such as cryo-electron microscopy and atomic force microscopy revealed that mimivirus have a vertex to vertex diameter of about 500 nm and an additional 100-200 nm of fibres on its surface yielding a virion of approximately 700 nm with icosahedral symmetry (Fig. 3) [37,38]. The inner nucleocapsid layer within which the genomic material is encompassed has a defined shape surrounded by an envelope and is separated from the capsid layer by a distance of about 30-50 nm [38]. The outermost capsid layer encoded by the major capsid protein L425 and has 31% amino acid similarity with the major capsid protein (Vp54)

of PBCV-1 [37]. Moreover, all mature mimivirus virions possessed a unique starfish-like structure at one of the five-fold vertices [39]. This structure was predicted to help the internal viral membrane to fuse with the host phagosome and thus serving the entry of viral DNA into the host [39].

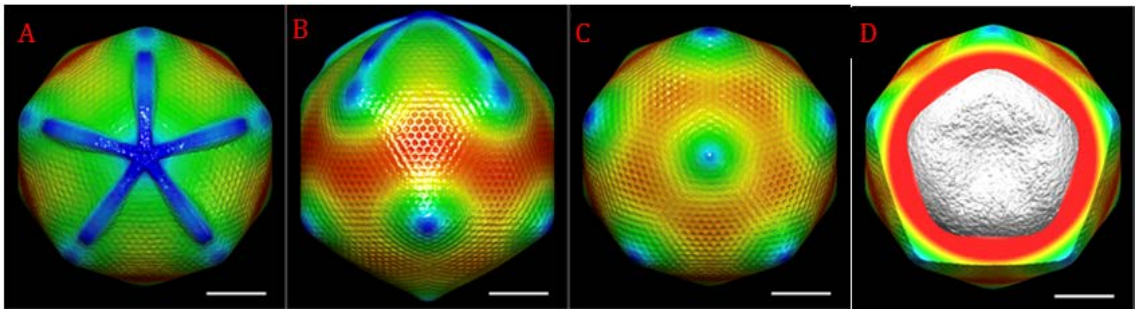


Fig. 3-Cryo electron microscopy reconstruction of mimivirus particles. (Figure adapted from [37]) (**A**) Top view of the starfish-like structure associated vertex. (**B, C**) Views from opposite sides of the vertex (**D**) The removal of the starfish-like structure to show the internal nucleocapsid and the icosahedral symmetry of the capsid. Scale bars= 100 nm. Colour code based on radial distance from the centre of the virus. Grey= 0-180 nm, red= 180-210 nm, rainbow colouring from red to blue= 210-250 nm.

The forest of dense fibres surrounding the entire surface of mimivirus was shown to be resistant to proteases except if they are first treated with lysozyme [37,38]. Cloudy material observed by atomic force microscopy surrounding the glycosylated fibres and the initial Gram-positive staining of mimivirus suggested the presence of peptidoglycans on its surface [1,37,38]. The biological roles of these glycosylated fibres have not been determined but a few studies reveal that it could act as a trap and attract amoeba [40]. Since mimivirus encodes for seven collagen-like genes, it was initially suggested that these fibres could be made up of collagen-like proteins [3]. However, the fibres were suggested to be not collagen-related since they were shown to be resistant to collagenases even after pre-treatment of virus particles by lysozyme [38].

Mimivirus replication

Viral replication was suggested to begin by phagocytosis of virus particles by the host amoebae (Fig. 4A) [41]. This is followed by the genomic DNA exit by the starfish vertex characterized on the virus structure which was called the ‘stargate’ a unique 5-fold structure which acts as a portal for genome delivery into the host cytoplasm [39]. It was

suggested that the stargate vertex operates as a device that helps the opening of the capsid protein which allows the outcropping of the inner viral membrane resulting in the fusion of viral-phagosome membrane [39]. Transmission electron microscopy and cryo-scanning electron microscopy revealed that viral replication and assembly takes place in cytoplasmic dense structures which are called virus factories from which new mature virions appear (Fig. 4B) [39,41,42].

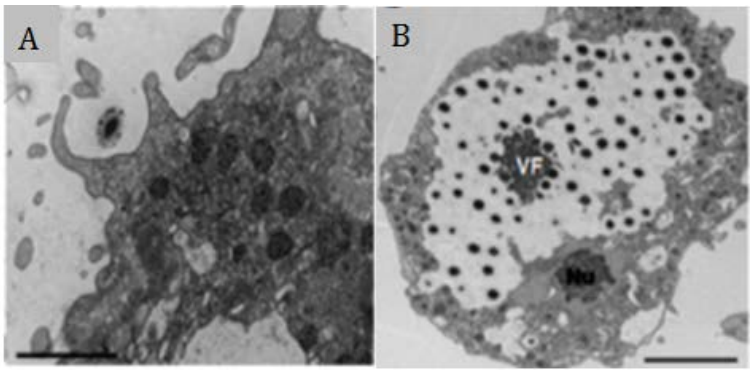


Fig. 4-Mimivirus replication in amoebae. (Figure adapted from [41]) (**A**) Transmission electron microscopy showing mimivirus being phagocytosed by an amoeba. Scale bar= 2 µm. (**B**) 12 hours post infection amoebae were filled with virus factories (VF) of new virions with a dense replication centre and the cell nucleus (Nu) was pushed to the periphery. Scale bar= 3 µm

Due to widespread replication, the expanding viral factories are not transported to the nucleus but fuse to form a large single factory [39,42]. Even though, the complete infection cycle seems to be cytoplasmic, there is a possibility that viral DNA enters the nucleus first probably for the first round of replication and then exits the nucleus to form the cytoplasmic virus factories [41]. Moreover, several host nuclear factors could escape or be actively delivered to the virus factories in the peri-nuclear space [42]. A graphic illustration of 24 h lytic mimivirus replication cycle is shown in Fig. 5.

Fluorescence microscopy and co-localization studies revealed that, along with replication, early transcription of messenger RNAs (mRNAs) occurs at separate sites in the cytoplasm which are adjacent to but are not co-localized with DNA replication sites [42]. This finding validates the presence of all five RNA polymerase subunits in the mimivirus core [43]. To conclude the replication and further propagation of such a giant virus composed of several viral proteins and RNA transcripts appears to be quite complex with various questions, such as host protein-organelle interaction and sources of nucleotides to synthesize such a giant DNA structure, still left unanswered [41].

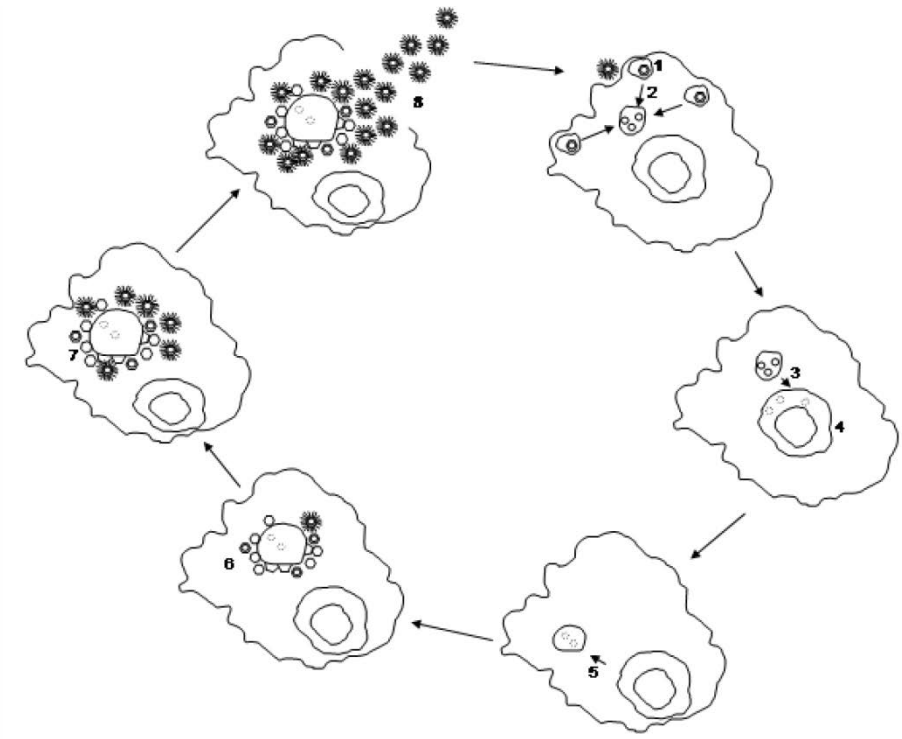


Fig. 5-Graphic illustration of mimivirus replication cycle. (Figure taken from [41]) 1-entry of mimivirus via a phagocytic vacuole, 2- fusion of phagocytic vacuoles, 3- mimivirus genomic material delivered to host cytoplasm, 4- probable first round of DNA replication in host nucleus, 5- formation of the first virus factory in the peri-nuclear space, 6-fusion of several virus factories into one large virus factory and assembly of proviral capsids, 7-mature virions floating around the virus factories in the cytoplasm, 8- mature virions with complete viral capsids surrounded by fibres released through cell lysis.

Ever increasing family of *Mimiviridae*

The isolation and discovery of mimivirus from a natural cooling water tower, led to the isolation and characterization of possible other giruses (giant viruses) [28] present in the environment through random sampling thus increasing the ever growing family of *Mimiviridae* of NCLDVs (Table 3). The current record holder of the largest viral genome is *Megavirus chilensis* isolated in 2011 from a marine environment off the coast in Chile [26]. The natural host of megavirus as such is still unknown, but is routinely cultured in another amoeba species *Acanthamoeba castellanii* [26]. Megavirus harbours a linear dsDNA genome with a size of 1,259,197 bp (~1.2 Mbp) within its icosahedral capsid surrounded by fibres [26]. The genome encodes for 1,120 putative proteins of which 594 orthologous proteins are shared between mimivirus and megavirus and 23% (258

proteins) have no mimivirus homologs [26,44]. In addition to the four aminoacyl-tRNA synthetases present in mimivirus, megavirus encodes for an additional three aminoacyl-tRNA synthetases which is suggestive of mimivirus-megavirus lineage evolving from a common ancestor via reductive evolution [26,44]. Structurally, mimivirus and megavirus virions show similar morphology of icosahedral capsid symmetry (~ 520 nm in diameter) however the fibres are strikingly shorter (~75 nm) than mimivirus (~120 nm) [26,44].

Table 3-Members of *Mimiviridae* family and its salient features (*Sputnik is not classified as a virus but as a virophage)

Virus	**Isolated from**	**Genome size (bp)**	**Capsid diameter (nm)**	**Surface fibres length (nm)**	**Putative protein coding genes**	**GenBank Accession number**
Megavirus	Marine coast in Chile	1,259,197	500-550	~75	1,120	NC_016072
Mamavirus	Cooling water tower in Paris	1,191,693	500-600	100-200	1,023	JF801956
Mimivirus	Cooling water tower in Bradford, UK	1,181,404	500-600	100-200	979	NC_014649
Moumouvirus	Cooling water tower in France	1,021,348	400-450	~100	930	NC_020104
CroV	Marine coast in Texas	~730,000	300	-	544	NC_014637
Sputnik*	Associated with mamavirus	18,343	50	-	21	NC_011132

Another giant virus isolated from a cooling water tower from Paris in 2008, which is effectively a strain of the original mimivirus, is called *A. castellanii* mamavirus [25,45]. The structural and cultural features of mamavirus closely resemble mimivirus encompassing a linear dsDNA genome of 1,191,693 bp which is 10,144 bp longer than mimivirus genome encoding for 1,023 putative proteins [3,25]. The genomes of mimivirus and mamavirus are very similar to each other except that mamavirus show exclusive characteristic divergence in the terminal regions especially a 13 kbp 5' terminal region present only in mamavirus [25]. A small 50 nm virus named Sputnik was identified within viral factories in the cytoplasm when amoeba cultures were co-infected with mimivirus and mamavirus [45]. The Sputnik has a circular dsDNA genome of 18,343 bp with 21 predicted protein coding genes [45]. Propagation of Sputnik virions impairs the growth of mimivirus particles itself in amoeba, suggesting it to be a candid parasite of viruses and hence classified as a virophage having similar functional analogies to bacteriophages [45].

After the megavirus, mamavirus and mimivirus, the next giant virus isolated from an industrial cooling water tower in south-east of France was named *A. polyphaga* moumouvirus [44]. The linear dsDNA genome of moumouvirus has 1,021,348 bp (200 kbp less than megavirus and 100 kbp less than mimivirus) encoding 930 putative proteins [44]. Even though it was isolated from a cooling water tower, moumouvirus is closely related to megavirus [26,44] which was isolated from a marine environment. Furthermore, recently the largest marine virus, isolated off the coast from Texas, infecting a marine zooplankton was identified *Cafeteria roenbergensis* virus (CroV) with a linear dsDNA genome of ~730 kbp of which the central 618 kbp region is sequenced [46]. The genome of CroV is encompassed in a large capsid of 300 nm diameter. The genome sequence of CroV revealed that it also belongs to the NCLDV group of viruses with mimivirus being its closest relative [46].

Origin and evolution of giant viruses

The genomic and proteomic analyses of mimivirus and other giant viruses presented novel features never identified in a virus before such as the genes encoding the

translation machinery, DNA repair enzymes, chaperones and other proteins involved in protein folding and enzymes involved in various metabolic pathways [3,25,43]. This led to many debateable questions regarding the evolution or co-evolution of viruses with its natural hosts, how did the virus acquire such complex genes and whether or not to include viruses in the tree of life. The natural hosts of mimivirus are free-living ubiquitously present amoebae, which have been called 'melting pot of evolution' as various micro-organisms have found ways to adapt and become resistant to digestion by amoeba and further propagate in vacuoles too [47,48]. These micro-organisms mainly consists of several bacteria belonging to wide range of prokaryotic phyla and are often pathogenic to humans including both obligatory and facultative intracellular organisms such as *Leginonella* species, *Parachlamydiaceae* species, *Rickettsiaceae* species and now viruses such as the mimivirus [47]. Phylogenetic analyses of mimivirus genes, showed the presence of many bacterial homologs indicating gene acquisition by horizontal gene transfer [49]. Since mimivirus shares the same environment with other bacteria, it was speculated that many genes were indeed acquired from live or degraded bacteria within amoebae via horizontal gene transfer [49-51]. Moreover, ~10% of mimivirus genome seems to be acquired from its natural amoeba host [49-51]. The presence of several intra-amoebal micro-organismic genomes further endorses that these amoebae houses different parasites or degraded bacteria at the same time leading to a 'melting pot of evolution' from various origins of life [47].

Considering the genome size of mimivirus of ~1.2 Mbp, it cannot be that the whole genome is acquired by horizontal gene transfer. Many phylogenetic studies have also revealed that mimivirus and other giant viruses might have undergone reductive evolution (i.e. continuous loss of a developing gene set) and have evolved or co-evolved with a more complex virus ancestor or co-existed with the last universal common ancestor [26,27]. One of the major reasons which led to the theory of reductive evolution is the presence of at least four aminoacyl-tRNA synthetases in mimivirus [3] and seven in megavirus [26] as a sign that these viruses must have evolved from a virus ancestor or last universal common ancestor or a cell encompassing a complete set of these synthetases [26,27]. Thus there seems to be a mix of both a complete history of multiple horizontal gene transfers and gene losses preceded and succeeded the appearance of an ancestor virus [4,5].

In view of these and many more arguments there are many who now believe that indeed viruses belong to the 'tree of life' and gives us in-depth insights into evolutionary picture of the current three domains of life-archaea, bacteria and eukarya [3,27,30,47,52]. However, also considering the inability of viruses to self-propagate, self-sustain and the cellular origin of their genes compels some researchers to believe that viruses do not belong to the 'tree of life' but however accepting that viruses do play an important role in understanding evolution of life [53].

Prevalence of giant viruses in environment

After the characterization of various giant viruses as mentioned above in *Acanthamoeba* species, it was important to identify if such giant viruses are present elsewhere in the environment as well. An initial study using the amoebal co-culture method isolated 19 giant viruses from 105 environmental samples coming from various sources such as lakes and rivers, cooling water towers, sea water, fountain water and even soil samples [54]. The sizes of the isolated viruses varied from 150-600 nm and many were identified to be closely related to Mimi/Mamavirus lineage [54].

Another study using high-throughput screening protocols identified 15 giant viruses from 1000 environmental samples from Tunisia coming from various ecological niches [55]. Of the giant viruses isolated, 11 were identified to be marseillevirus-like and 4 mimivirus-like [55]. Furthermore, the identification of a large virus CroV from a marine environment and further metagenomic studies suggests the presence of many large and giant viruses of eukaryotes present in diverse marine environments as well [29]. These various studies clearly shows us that many *Acanthamoeba*-growing and other various giant viruses are distributed and routinely found in the environment in varied ecosystems suggesting that humans could be at potential risk of coming in contact with giant viruses.

Potential pathogenicity of mimivirus in humans

There have been several case studies to identify the presence of antibodies specific to mimivirus in the human population. Amoeba-associated micro-organisms including

mimivirus have been postulated to be one of the causative agents of community and hospital acquired pneumonia [2,56,57]. Seroconversion has been observed in many community and hospital acquired pneumonia patients [2,56,57]. A clinical case study revealed that a technician working routinely with mimivirus was seroconverted in a case of severe pneumonia [58]. Furthermore, mimivirus DNA was identified in the bronchoalveolar lavage fluid of only one pneumonia patient in Canada [57]. Likewise, in a small *in vivo* mouse model, it was shown that mice inoculated with mimivirus through an intracardiac route, developed acute atypical pneumonia and mimivirus could be re-cultured from lung tissues isolated from these mice [59]. However, the seroconversion to mimivirus in pneumonia patients is quite debatable and controversial as in many studies, they failed to isolate mimivirus from pneumonia patients samples suggesting that mimivirus is not a common respiratory pathogen [60,61]. Very recently, using metagenomic studies mimivirus and marseillevirus like sequences were identified in gut microbiota of a young man from Senegal and identification of a new giant virus in human stool samples named Senegalvirus belonging to putative *Marseilleviridae* family, suggesting the possible presence of giant viruses in human gut microbiota [62]. Even though in a few studies pneumonia patients were not associated with mimivirus [60,61], a few studies did link the potential of mimivirus in pneumonia [2,56-58] and also the presence of giant viruses in human stool samples [62], the probable pathogenicity of mimivirus in humans should not be avoided completely and could be considered as an emerging potential human pathogen .

Furthermore, using an *in vitro* system it was clearly demonstrated that mimivirus were taken up by professional macrophages only by the process of phagocytosis and not by non-professional macrophages which led to an active replication of mimivirus within macrophages [63]. Phagocytosis is a process used against various bacteria and parasites by their hosts but this was the first indication that it can also be used by a virus [48,63,64].This study suggested that macrophages could be a potential human target for mimivirus and that it propagates in alveolar macrophages causing human and murine pneumonia [59,63].

(Dis)Similarities between human and mimivirus collagens

Collagens or collagen-like proteins are found in all metazoans including flies, worms and sponges, playing a vital role not only in structural functions but also in chemotaxis, growth and differentiation, movement, cell to cell adhesion, wound healing and much more [65-67]. Collagens in animals are the most abundant proteins found in the extracellular matrix playing an important role in retaining the various tissue structures and physiological functions [65-67]. In humans, collagen represents about one third of all the protein mass. Humans have at least 28 different collagen types encoded by more than 40 genes [66,67]. Considering the different supramolecular organization and structure, collagens are classified into many groups and are widely distributed in various tissues (Table 4). Fibril-forming collagens constitute about 90% of the total collagen within the various classes of collagens found in humans [65]. Defects or mutations in collagen genes can lead to serious disorders such as osteogenesis imperfecta (defective connective tissue), chondrodysplasias (arrested development and bone deformities), Ehler-Danlos syndrome (defective connective tissue) and many more [65-67].

Table 4-Various collagen types in humans with its classification and widespread tissue distribution (table adapted from [65])

Collagen type	**Collagen classification**	**Tissue distribution**
I	Fibril-forming collagens	Bone, dermis, tendon, ligaments, cornea
II	Fibril-forming collagens	Cartilage
III	Fibril-forming collagens	Skin, vessel walls, reticular fibres of various tissues
V	Fibril-forming collagens	Lung, cornea, bone, fetal membranes
IV	Basement-membrane collagens	Basement membranes
VI	Microfibrillar collagens	Dermis, cartilage, placenta, lungs, vessel wall, intervertebral disc
VII	Anchoring fibrils	Skin, epidermal junctions, oral mucosa
XII	Fibril-associated collagens	Ligaments, tendons

XIV	Fibril-associated collagens	Dermis, tendon, vessel wall, placenta, lungs, liver
XX	Fibril-associated collagens	Corneal epithelium, embryonic skin, sternal cartilage
XXI	Fibril-associated collagens	Blood vessel wall
XIII	Transmembrane collagens	Epidermis, hair follicle, intestine, chondrocytes
XVII	Transmembrane collagens	Epidermal junctions

In general, the structure of collagens consist of three polypeptide chains (homo or heterotrimer) coiled into a left-handed helix which further wounds around its common axis to form a triple helix and is synthesized in the ER and transported through Golgi in large transport vesicles (Fig. 6) [67]. Depending upon the collagen type, the length of the triple helix varies considerably. At least one of the polypeptide chains are characterized by the repeating triple G-X-Y repeats, where G is glycine; X and Y can be any amino acid other than glycine but are most often proline and lysine [66]. Glycine present at every third position is necessary for the packing of the helical configuration of collagens [67]. Collagens can be post-translationally modified by hydroxylation of selected proline and lysine residues [68,69]. The amount of 4-hydroxyproline residues is essential in forming intra-molecular hydrogen bonds and helps in the thermal stability of collagens [70]. Further, hydroxylysine can be glycosylated by addition of galactose (in β1,O-linkage) and then glucose (in α1,2-linkage) and this glycosylation pattern is conserved from sponges to humans [71,72]. The enzyme catalysing the addition of galactose to hydroxylysine residues in humans have been identified to be a galactosyltransferase (GLT25D1/D2) [73] but however the enzyme further glucosylating this galactose is not yet identified. After the biosynthesis and post-translational modifications of collagens in the ER and passing through the Golgi, it is passed through the secretory pathway out in large transport vesicles wherein specific proteases cleave the N and C propeptides [65-67]. Further, single chains spontaneously self-assemble and forms covalent crosslinks initiated by oxidation of hydroxylysine residues [65-67].

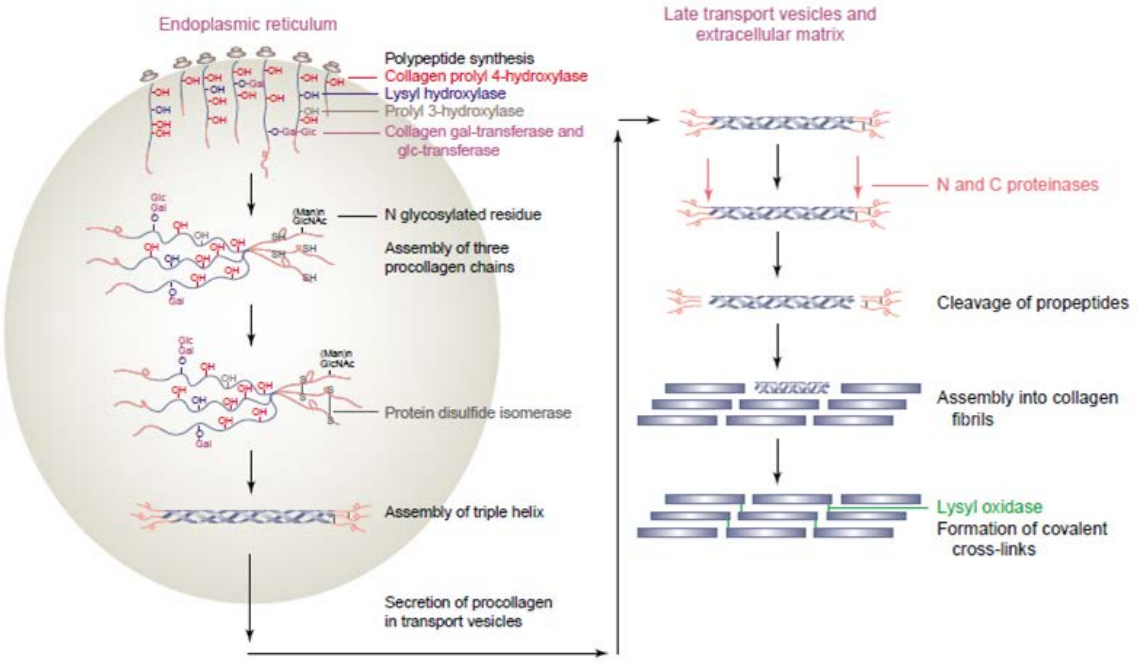


Fig. 6-Major steps in biosynthesis of fibril-forming collagens. (Figure taken from [67]) The single polypeptide chains of collagens are synthesized by membrane bound ribosomes and are secreted in the lumen of the endoplasmic reticulum where the following major steps occurs (1) hydroxylation of certain prolines and lysines by their respective hydroxylases (2) glycosylation of a few hydroxylysines by addition of galactose and/or glucose and certain aspargine residue are N-glycosylated (3) association of three C-propeptides directed by recognition sequences (4) formation of intra- and inter-molecular disulphide bonds (5) assembly of the triple helix from C-terminal towards the N-terminal in a zipper-like fashion (6) procollagen molecules are transported through Golgi in the transport vesicles (7) successive steps are cleaving of N- and C-propeptides by specific proteinases followed by self-assembly into fibrils and covalent cross-linking of certain lysine and hydroxylysine residues.

Interestingly, many studies have revealed that some bacteria and bacteriophages also code for collagen-like regions in their genomes [74]. Collagen-like proteins are also expressed in some fungi such as *Metarhizium anisopliae* [75] and bacteria such as *Streptococcus pyogenes* [76]. Also, some viruses encode for collagen-like proteins such as the shrimp white spot syndrome virus [77], lymphocystis disease virus [78] and giant viruses such as mimivirus [3]. The genome of mimivirus revealed 7 ORFs which encoded for collagen-like proteins (Table 5). Furthermore, genome sequencing of other giant viruses in the *Mimiviridae* family revealed the presence of several genes encoding for collagen-like proteins; megavirus has 12 (GenBank Accession number-NC_016072), mamavirus has 7 (GenBank Accession number-JF801956), moumouvirus has 9 (GenBank Accession number-NC_020104) and remarkably even the small virophage Sputnik has 2 collagen-like proteins encoded by its small genome (GenBank Accession

number-NC_011132). The functions of such heavy molecular weight proteins like collagens in giant viruses is yet unknown. More interestingly these viral collagen-like proteins also show repeating G-X-Y repeats as seen in human/animal collagens but however in contrast have a poor proline content (Fig. 7).

Table 5-Collagen-like proteins in mimivirus genome (GenBank Accession number-NC_014649)

Mimivirus ORF	**Annotation**	**Predicted size of annotated protein**	**Nucleotide length of annotated ORF**
L71	Collagen-like protein 1	94 kDa	2932 bp
R196	Collagen-like protein 2	158 kDa	4902 bp
R239	Collagen-like protein 3	91 kDa	2892 bp
R240	Collagen-like protein 4	80 kDa	2460 bp
R241	Collagen-like protein 5	78 kDa	2443 bp
L668	Collagen-like protein 6	141 kDa	4220 bp
L669	Collagen-like protein 7	194 kDa	5814 bp

MSRITCPITDCKCKCNKNNCVYCVMGRQGLPGPKGSSGNSIYVGTGVPSPFLGNNGDLYI
DSSTGLLYAKVNGVWVPQGSLKGDPGASGSKGEKGDKGSSGEAGLKGEQGTKGEQGDQGE
QGDKGDKGDKGDVGAKGDQGDKGDQGDVGAKGDQGDKGDQGDVGAKGDQGDKGDKGDQGD
KGDVGDPGVKGDKGDTGDKGDKGDKGDKGQNGSEILFGLGIPSPDLGEDGDVYIDTLTGN
VYQKIGGVWVLETNIKGEKGDQGDKGDTGSKGDQGDKGDQGDKGDQGDKGDVGDKGNKGD
TGSKGDVGDKGDVGDKGDKGDTGDKGDKGDTGDKGDKGDVGDKGDKGDVGDKGDVGDKGD
VGDKGDKGDTGDKGDKGDIGDKGDKGDIGDKGDKGDIGDKGDKGDVGDKGDKGDKGDIGD
KGDKGDIGDKGDKGDKGDKGENGSGILFGLGIPSPDLGEDGDIYIDTLTGNVYQKIGGVW
VLETSIKGEKGDKGDTGDKGDTGDKGDTGDKGDTGDKGDTGDKGDVGDKGDVGDKGDVGD
KGDVGDKGDKGDIGDKGDKGDLGDKGDKGDVGDKGDVGDKGDKGDIGDKGDKGDLGDKGD
KGDVGDKGDKGDVGDKGDKGDIGDKGDKGDVGDKGDKGDIGDKGDKGDKGDVGSKGDKGD
KGDVGDKGDKGDVGSKGDKGDKGDKGDVGPVGASILFGAGVPSPTTGENGDSYIDNSTGV
FYLKINDVWVPQTNIKGDKGDKGDKGDKGDKGDTGDVGLKGDTGTPGSGPIIPYSSGLTP
VALAVVAVAGGGIADTGASYDFGVSSPSVTLVGVNLDFTGPVQGLLPNMAWSAPRDTVIT
SLATAFQVSVAISAVLEPIFLRTQVYRELAANPGVFEPLAGAIVEFDVASSALISVGTVF
RGIVTGLSIPVNAGDRLIVFANTRTTSLISVGTVTGFISSGLALA

Fig. 7-Protein sequence of mimivirus collagen-like protein 1 L71. Here as an example of mimivirus collagens, the protein sequence of L71 is shown which elucidates the continuous presence of G-X-Y triple repeats (marked in red) as is also observed in human/animal collagens as well.

Glycosylation machinery of mimivirus: Like PBCV-1, mimivirus also encodes its own glycosylation machinery responsible for synthesis of sugars (di, oligo and polysaccharides) which are involved in post-translational modifications (N- and O-glycosylation) of its proteins such as its own capsid protein encoded by L425 [3]. As mentioned above, the first virally encoded glycosyltransferase and N-glycosylation pathway have been identified in the PBCV-1 [17,20] and genome sequence analysis of mimivirus indicated many genes potentially involved in glycosylation of its structural proteins and sugar modifying enzymes. Mimivirus encodes for a complete nucleotide sugar metabolite pathway (two enzymes) for biosynthesis of sugars such as rhamnose in contrast to PBCV-1 wherein only one enzyme is encoded by the virus and depends on the host for the second enzyme to complete the pathway [17,79]. This virus encoded pathway and enzyme activity to synthesize rhamnose is similar to as identified in plants [79]. Furthermore, mimivirus has the potential of infecting mammalian macrophages through phagocytosis [63], and since mammalian cells lack the biosynthetic pathway for rhamnose, a full virus encoded pathway provides the virus the essential sugar for its own protein post-translation modification in its host [79]. Furthermore, sugar analyses of mimivirus revealed the presence of rhamnose, glucose, N-acetylglucosamine and an unusual monosaccharide called viosamine [80]. Interestingly, this monosaccharide viosamine is also typically found in many bacteria such as in lipopolysaccharide antigens of *Shigella disenteriae* [81], Shiga toxin producing *Escherichia coli* [81], exosporium of *Bacillus anthracis* [82] and flagellins of *Psuedomonas syringae* [83]. The expression of these enzymes occurs during the late stages of mimivirus replication suggesting a possible link between virion formation and addition of glycosylated fibres on mature virions [79,80]. These studies clearly suggest that mimivirus also encodes for host-independent glycosylation machinery [79,80].

Moreover, recently it has been shown that mimivirus encodes for a novel bi-functional hydroxylase and glycosyltransferase enzyme (L230) which hyroxylates the lysine residues on collagen like proteins and further glycosylates it [84]. The seven mimivirus collagen-like proteins are expressed during the viral life cycle along with this novel bi-functional enzyme (L230) capable of hydroxylating lysine residues and glycosylating it further [84]. This bi-functional enzyme interestingly adds glucose directly to the hydroxylysine residues and is in contrast to the conserved glycosylation pattern of collagens of galactose-glucose [71,72,84].This suggests that possibly mimivirus

collagens can also be post-translationally modified and that too in a unique form. Thus, this variability in human and mimivirus collagen glycosylation could lead to antigen or molecular mimicry (Fig. 8).

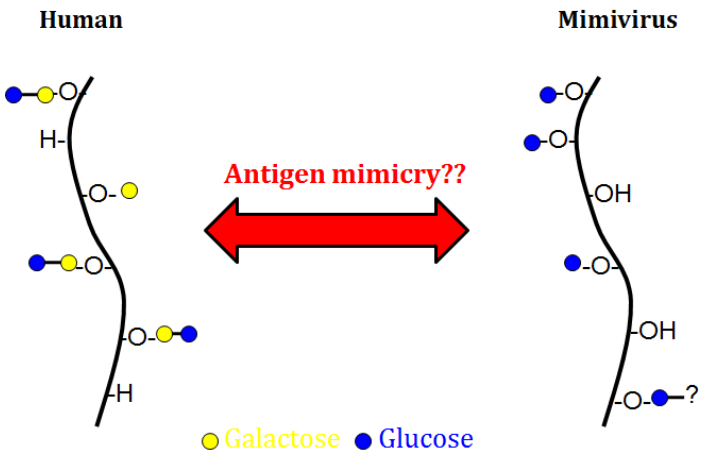


Fig. 8-Variablity in collagen glycosylation between humans and mimivirus. (Figure credits-Dr. K. Luther) The figure elucidates the conserved glycosylation pattern in human collagens whereby the hydroxylysine residues are first galactosylated (yellow circles) and then glucosylated (blue circles). However, in mimivirus collagens the hydroxylysine residues are directly first glucosylated and it is yet unknown if any other sugar is further added to the glucose. This variability in collagen glycosylation could lead to antigen mimicry.

Additionally, it has been reported in our lab that at least one of the mimivirus collagens, L71, is expressed on the surface of the virus particles (Shah *et. al.*, 2013, appendix). The L71 protein has 945 amino acids with 4 collagen domains encompassing 561 amino acids (Fig. 9). Furthermore, L71 also possessed a stretch of amino acids with 73% sequence identity to a major collagen type II T cell epitope which is identified in an autoimmune disease like rheumatoid arthritis [85,86]. Interestingly, the linker sequences between the collagenous domain of L71 show a high percentage of sequence identity (Fig. 7 and 9) and this is maintained even in other collagen-like proteins in mimivirus (R196, L668, L669) and is also conserved even in other collagen-like proteins of giant viruses such as mamavirus, megavirus and moumouvirus.

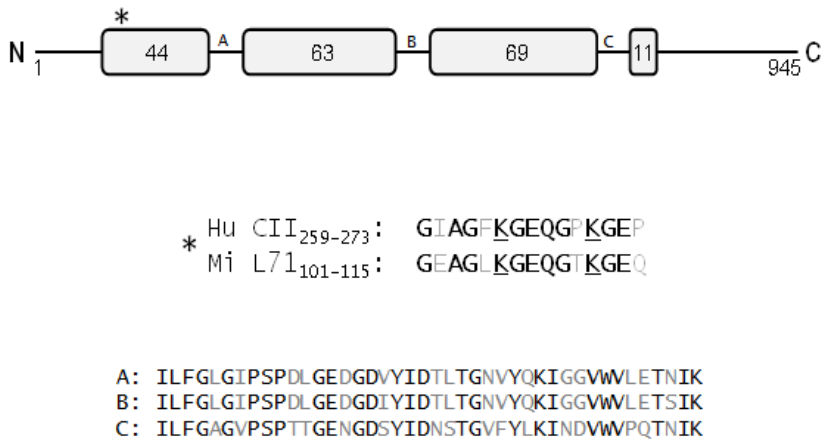


Fig. 9-Domain organization of mimivirus collagen-like protein 1 L71. (Figure adapted from Shah *et. al.*, 2013) The four collagen domains of L71 are shown as grey boxes with the number of G-X-Y repeats given inside. The asterisk shows the position of the sequence motif similar to the epitope human collagen type II recognized as immunodominant in rheumatoid arthritis [85]. The sequence of this human collagen type II (Hu CII) epitope encompassing amino acids 259-273 is shown aligned with the corresponding sequence of mimivirus L71 (Mi L71) encompassing amino acids 101-115. Furthermore, the linker sequences (A, B, C) between the collagenous domains also share sequence similarity.

Thus, the expression of several collagens by mimivirus, the variability in its glycosylation and also the presence of similar amino acid sequence to major human T cell epitope identified in human autoimmune disease of rheumatoid arthritis led us to hypothesize that whether or not mimivirus has the potential in triggering an autoimmune response against animal collagens and play a role in inducing autoimmunity to collagens.

Rheumatoid Arthritis

Introduction to autoimmunity

It was in the beginning of the twentieth century that Nobel Laureate Paul Ehrlich predicted the concept of autoimmunity and termed it as "*horror autotoxicus*" [87]. Generally the body's immune system can recognize what is 'foreign' and mount an appropriate immune response against it. However under certain circumstances, the immune system fails to distinguish between 'foreign' and 'self' and mounts an immune response against self-proteins and self-target organs. Autoimmune reactions can prove to cause fatal damages to the body with severe damage to cells, tissues and organs. Autoimmune diseases may result from different interacting components such as genetic, environmental and immune(dys)regulatory (Fig. 10).

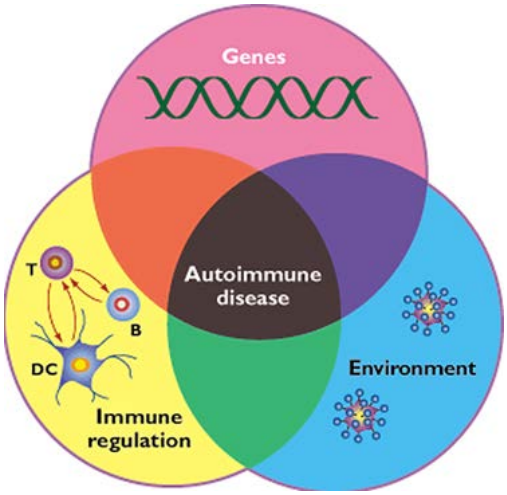


Fig. 10-Requirements for the development of autoimmune diseases. (Figure taken from [88]) Autoimmune diseases are a multifactorial process requiring certain genetic risk variants along with certain environmental factors such as infection with pathogens along with immunoregulatory mechanisms.

Immune tolerance

The immune system is referred to as the body's sixth sense [89]. It can fight any possible foreign protein or microorganism to protect us from further phenotypic and clinical complications [89]. It has developed several mechanisms to overcome and protect from reacting to self-proteins and is called 'tolerance' (state of unresponsiveness to an antigen). Tolerance is generally antigen specific and induction

of tolerance does not render the immune system inactive but is specific to the tolerogenic antigen [89]. It is when lymphocytes escape the tolerance mechanisms of the body and are activated, autoimmune reactions are developed [90]. Two main tolerance mechanisms exist- central and peripheral tolerance (Fig. 11).

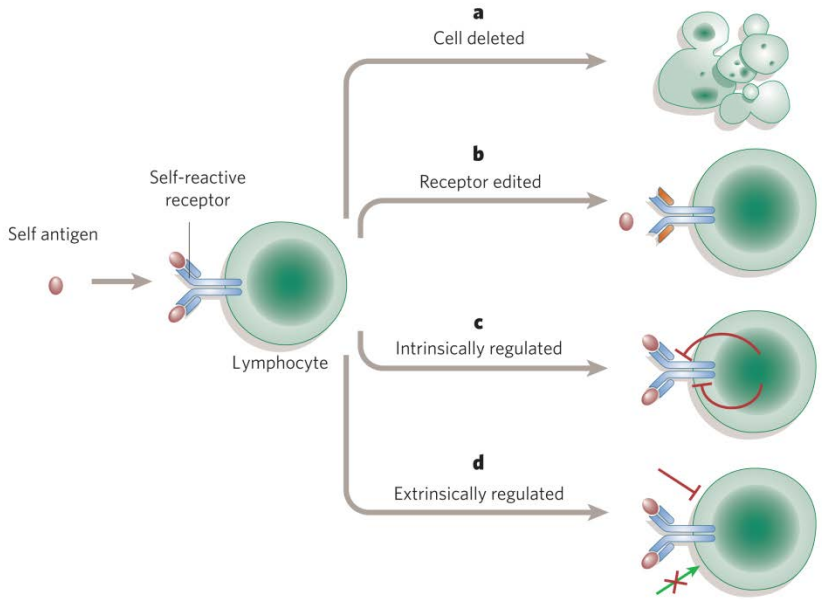


Fig. 11-Cellular strategies adopted by B and T lymphocytes to regulate auto-reactive receptors. (Figure adapted from [89]) (**a**) Self-reactive lymphocytes are deleted through apoptosis (**b**) Receptor is edited to less self-reactive (**c**) Certain intrinsic biochemical and/or gene expression changes the self-reactive receptor's ability to activate the cell (**d**) Limiting the co-stimulatory signals and cytokine production extrinsically regulates the self-reactive lymphocytes**.**

Central tolerance

The lymphocytes mature in their respective lymphoid organs- the bone marrow for B cells and thymus for T cells. The main mechanism to induce tolerance is the early detection and deletion of immature lymphocytes through apoptosis that may recognize self-antigens in the lymphoid organs where they mature and is termed as central tolerance [89,90]. This is mainly coordinated by the receptors expressed on the lymphocytes- the immunoglobulins (Ig) as B-cell receptor (BCR) and the T-cell receptors (TCR) presented on T cells. Various genetic recombinations in the variable

(V), diversity (D) and joining (J) gene segments gives rise to functional BCR and TCR [91,92]. If the immune system recognizes the potential of these receptors to be auto-reactive then it deletes them by apoptosis, what is also known as negative selection of lymphocytes (Fig. 11a) [89,91,93,94]. Furthermore, some developing B cells can undergo a process of receptor editing, via V(D)J recombination, in which the auto-antigen specific V region is edited and switched to a different V region rendering it inactive for autoimmunity (Fig. 11b) [92]. Thus mainly through the process of negative selection of lymphocytes and receptor editing auto-reactive B cells can be eliminated in the bone marrow. T cells also undergo central tolerance in thymus also by the process of apoptosis (i.e. negative selection) by those cells which show low or high affinity to self-antigen presented by the antigen presenting cells (APCs) when the peptide is presented with a specific type of major histocompatibility complex (MHC) molecules (Fig. 12) [94]. However, those T cells which show intermediate affinity, mature and leave the thymus enter the periphery where peripheral tolerance tries to eliminate them [94]. Thus in general, central tolerance limits the development of potential auto-reactive B and T cells.

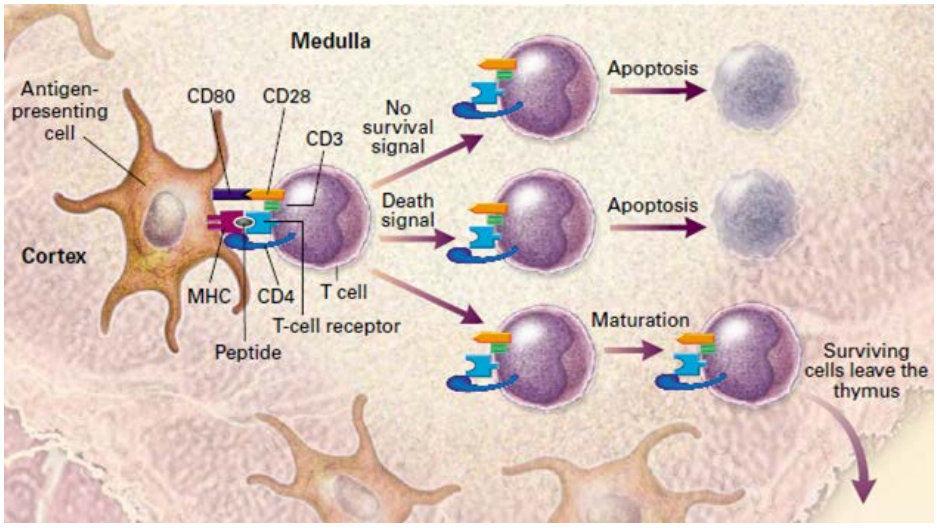


Fig. 12-Central tolerance mechanisms for induction of immune tolerance. (Figure taken from [94]) immature T cells educating themselves in the cortex and medulla of the thymus, encounter novel antigen presented by antigen presenting cell. Cells displaying TCRs having low or high affinity to the self-peptide and MHC molecule receives a signal to switch off and undergo spontaneous apoptosis. The remaining T cells showing intermediate affinity to these complexes can mature and leave the thymus and enter the periphery.

Peripheral tolerance

If lymphocytes escape the deletion process of central tolerance, a second and more stringent mechanism known as peripheral tolerance exists in the immune system. Central tolerance is not always a complete fool proof process since it is not necessary that in the central lymphoid organs all self-antigens are expressed and furthermore a minimum threshold is required for affinity to self-antigens before they are deleted allowing self-reactive lymphocytes to escape the sorting-out process [94]. In the periphery, self-reactive lymphocytes can be regulated intrinsically or extrinsically. BCR and TCR can be downregulated via a series of intrinsic biochemical changes on the cells which express them [89]. A state of clonal anergy (i.e. complete unresponsiveness to a stimulus) can be achieved by B cells expressing self-reactive BCR [95,96]. Certain important signalling cascades such as tyrosine kinase signalling pathway can be downregulated in the presence of weak stimulation from various factors such as nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) can render self-reactive B cells anergic (Fig. 11c) [97,98]. T cells can also undergo immunological ignorance wherein T cells ignore the self-antigen or can be physically concealed by an anatomical barrier such as the blood-brain barrier (Fig. 13) [94]. T cells can also be deleted by interaction between Fas and its respective Fas ligand which induces apoptosis (Fig. 13) [99]. Self-reactive T cells can be rendered anergic in the presence of certain inhibitory receptors such as cytotoxic T-lymphocyte antigen 4 (CTLA4) which competes with CD28 expressed on APCs and binds to B7 and transmits inhibitory signals for T cell activation (Fig. 13) [100-102]. Furthermore, it has been shown that absence of CTLA4 (also known as CD152) can cause substantial build-up of self-reactive T cells in lymphoid tissues [102,103]. In addition, a new subset of T cells ($CD4^+$ $CD25^+$) called regulatory T cells (T_{reg} cells) are produced in the thymus which show intermediate affinity to self-antigens, seem to play a role in inducing peripheral tolerance as well [104]. Lymphocytes can be extrinsically regulated by the production of cytokines such as interleukin (IL) 10 or transforming growth factor beta (TGFβ) which is in fact regulated by T_{reg} cells actively suppressing the immune response (Fig. 11d and 13) [89,94,104]. One of the co-stimulatory signal between CD40 ligand (CD40L) a surface protein of T cells and its secreted cytokines such as IL 2, IL 5, IL 21 and B cells are

required for production of antibodies by B cells and thus limiting these co-stimulatory signals also help in extrinsically providing tolerance [105,106].

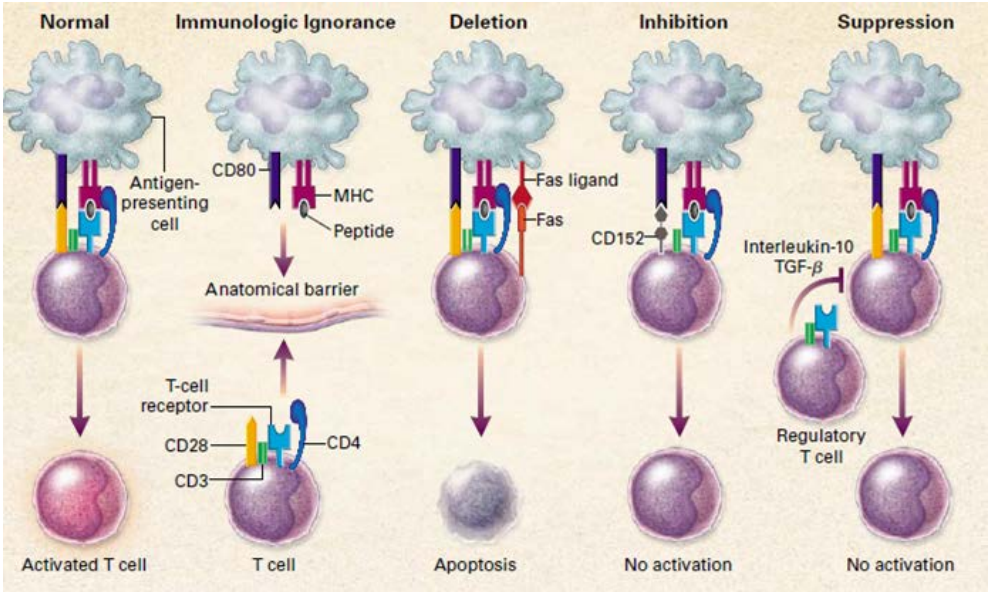


Fig. 13-Peripheral tolerance mechanisms for induction of immune tolerance by T cells. (Figure taken from [94]) T cells can be physically concealed such as the blood-brain barrier and thus cannot become activated sometimes referred to as immunological ignorance. Certain T cells express the Fas molecule (CD95) and receive apoptotic signals from the Fas ligand expressed by self-reactive antigen presenting cells. Inhibitory receptors on self-reactive T cells such as CD152 (also known as CTLA4) render them anergic and thus are not activated. Regulatory T cells suppress self-reactive T cells in the presence of certain cytokines such as TGFβ and IL 10.

Breakdown of immune tolerance

The failure of host's tolerance mechanisms to protect the host from self-reactive lymphocytes leads to innumerable autoimmune diseases. One of the reasons postulated as to how tolerance is broken, is the fact that deletions of intrathymic auto-reactive T cells have individual variations [107]. For example, the gene known to contribute to the susceptibility of diabetes mellitus is also known to determine intrathymic levels of (pro)insulin especially during development and childhood which may help in promoting negative selection of insulin specific T cells conferring autoimmunity to insulin causing diabetes [108,109]. Certain foreign agents termed as haptens are known to render auto-antigens immunogenic. For example, certain drugs used in haemolytic anemia such as penicillin and cephalosporin have the capacity to bind to red blood cell membrane and create neo-antigens provoking auto-antibody responses causing anemia [110]. The

synthesis of wide variety of glycosylated proteins differ between the thymus and peripheral tissues and thus there is a possibility that glycoproteins may be novel in the periphery and depend solely on the peripheral tolerance mechanisms, thus influencing tolerance [111].

Furthermore, various mechanisms are proposed because of environmental factors such as infections with certain microbes linking the breakage of tolerance and causing autoimmune diseases. Due to certain tissue damages caused by infections, there is a possibility that sequestered auto-antigens are released and hence potentially self-reactive T cells are activated [112]. Certain exotoxins secreted by microbes may act as 'super-antigens' and has the capability for activating up to 20% of the normal repertoire of T cell population [94,113]. Presentation of self-antigens due to the induction of pro-inflammatory cytokines and co-stimulatory molecules by infectious microbes also help in breaking immune tolerance [114]. Another important concern due to inflammation is the up regulation of certain glycosyltransferases and thus altering the immunogenicity of self-proteins [115]. Furthermore, whenever there is a strong local activation of APCs, a phenomenon called 'epitope spreading' occurs wherein antigens can be over presented or over processed thus priming large numbers of T cells, potentially against self-antigens [116]. It was also shown that when lymphotrophic viruses infect B cells, polyclonal activation occurs [117]. This results in B cell proliferation leading to building up of circulating immune complexes [117]. Alternatively, structural similarities between microbial and self-antigens have been also linked to the pathogenesis of many autoimmune diseases by a phenomenon termed as molecular or antigen mimicry by breaking immune tolerance. According to the phenomenon of molecular mimicry, if a host comes in contact with an infectious agent that has immunologically similar antigens to the host, then the host mounts an immune response on foreign antigens [118]. But since due to the similar structure or amino acid sequence of the foreign antigenic molecules, it mimics the host molecules and the immune response cross-reacts and tolerance to auto-antigens is broken because of which autoimmunity is established leading to severe tissue and/or organ damage [118]. In many autoimmune diseases pathogenesis has been linked to the phenomenon of molecular mimicry. In infections with *Campylobacter jejuni* which generally causes gastroenteritis, antibodies recognizing *Campylobacter* lipooligosaccharides can cross-react with GM1 gangliosides

on nerve cells leading to Guillain-Barré syndrome [119,120]. In infections with lymphocytic chriomeningitis virus, antibodies against the viral proteins cross-react with host's central nervous system tissues containing myelin basic protein causing demyelination and motor dysfunction leading to multiple sclerosis [121,122]. However, it is difficult to confirm that a particular autoimmune disease is caused solely by an infectious agent. It could be that before the disease is clinically manifested the infectious process has been determined long before or the infection itself could be hidden making molecular mimicry an epi-phenomenon but the possibility cannot be ruled out [118]. Thus it seems that the clinically manifested autoimmune diseases are only the 'tip of the iceberg' since many biological processes are involved to evade immune tolerance before any autoimmune diseases are developed (Fig. 14) [123].

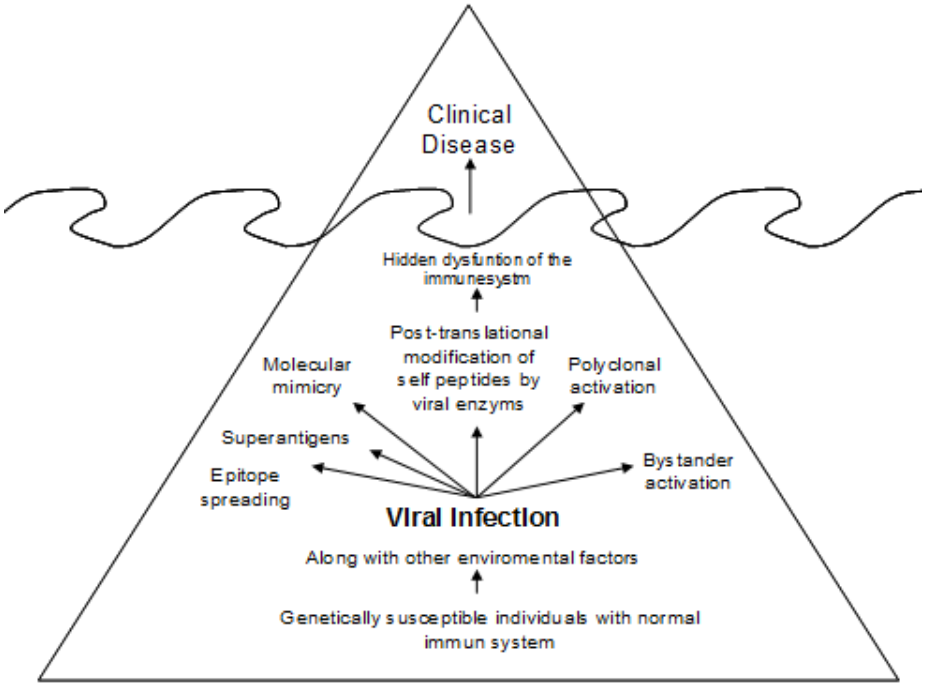


Fig. 14-'Tip of the iceberg' model proposed for pathogenesis of autoimmune diseases. (Figure taken from [123]) As this model proposes, there can be several events taking place before autoimmune diseases are triggered. For example, an individual born with genetic susceptibility to an autoimmune disease may encounter numerous environmental factors such as viral infections. This would in turn lead to a dysfunctional immune system due to the several mechanisms such as polyclonal activation of lymphocytes, molecular mimicry, epitope spreading, differential post-translational modifications of host proteins, etc. eventually resulting in clinical manifestation of the autoimmune disease.

Some of the common autoimmune diseases occurring in humans leading to serious fatal consequences and are widely studied are thyroiditis in which autoantibodies towards specific thyroid organ cells are produced which causes severe inflammation of the thyroid gland [124]; insulin dependent diabetes mellitus wherein autoimmune attack destroys the insulin producing beta cells of the islets of Langerhans in pancreas resulting in increased levels of blood glucose [125]; systemic lupus erythematosus caused by the production of autoantibodies towards a wide range of tissue antigens such as erythrocytes, platelets, leukocytes and skin tissue causing sometimes fatal haemolytic anemia [126]; multiple sclerosis caused by production of auto-reactive T cells that target the components of central nervous system destroying the myelin sheet insulating the nerve fibres causing neurologic dysfunctions [121]; rheumatoid arthritis a systemic inflammatory autoimmune disease primarily affecting joint tissue [127]. In the context of this thesis, causes, pathogenesis, immune regulations and possible treatments towards rheumatoid arthritis will be explained in detail in the following sections.

Causes of rheumatoid arthritis

Rheumatoid arthritis is a systemic inflammatory autoimmune disease primarily targeting articular joint tissue affecting 0.5-1% of world population [128]. This systemic disease results in chronic inflammation of synovial joints causing severe damage to the cartilage and bone structure which further leads to various clinically important potential comorbidities as a collective inflammatory response (Fig. 15) [127-129].

One of the major clinical markers for rheumatoid arthritis is the autoantibody called rheumatoid factor, directed against the Fc domain of IgG molecules and is one of the major classification criteria for rheumatoid arthritis set by the American College of Rheumatology [127-129]. However, the presence of rheumatoid factor is not restricted to rheumatoid arthritis patients only but is also observed in elderly people without the disease as well [127-129]. Another antibody response recognized in rheumatoid arthritis, targets the presence of antibodies to citrullinated protein antigen which is sometimes referred to as anti-CCP (cyclic citrullinated peptide) [127-129]. Under the

influence of some genetic and environmental factors, an enzyme peptidylarginine deiminases post-translationally modifies amino acid arginine to citrulline (process called citrullination) [127-129].

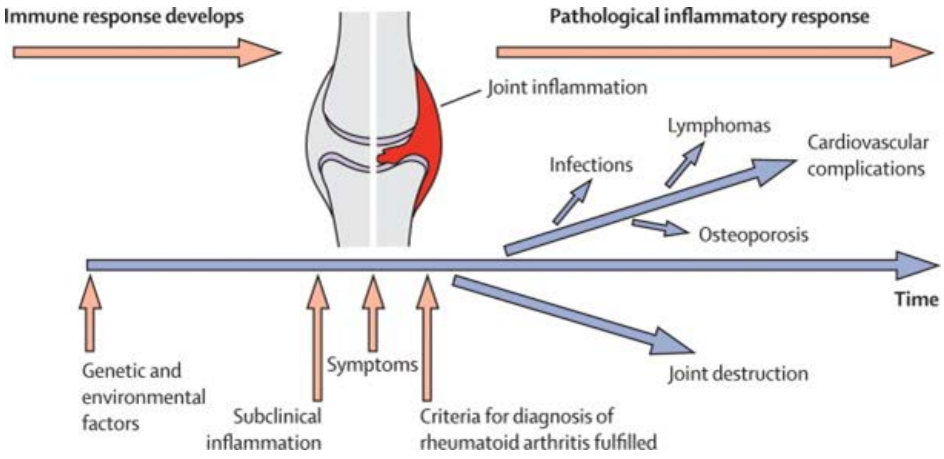


Fig. 15-Longitudinal course of pathogenesis of rheumatoid arthritis. (Figure taken from [127]) The figure shows that in the course of time, in genetically susceptible individuals under the influence of some environmental factors the immune response is triggered resulting in joint inflammation. Moreover, if inflammation becomes chronic it fulfils the criteria for rheumatoid arthritis and along with inflammation it leads to complex pathological comorbidities as a collective inflammatory response.

Considering the high specificity of antibodies to citrullinated protein antigen during rheumatoid arthritis than rheumatoid factor it is widely used as a diagnostic tool for detection of early disease [127-129]. Many patients being positive for antibodies to citrullinated protein antigen and rheumatoid factor have also conspicuous manifestations of joint destructions along with other comorbidities such as cardiovascular disease and other extra-articular signs [127]. As with every autoimmune disease, rheumatoid arthritis too is a complex disease wherein several genetic, environmental and certain stochastic (random/chance) factors act in concert to clinically manifest the disease [127].

Genetic factors

It was in the 1970s that the genetic association of rheumatoid arthritis with HLA alleles were described and investigated. A common amino acid sequence (QKRAA, QQRAA,

KKRAA) termed as the 'shared epitope' present in the β-chain of HLA-DR molecule responsible for the peptide binding grove confers susceptibility to rheumatoid arthritis [130-132]. The shared epitope alleles is generally expressed on the HLA-DRB1*01, DRB1*04 and DRB1*10 and are high risk variants of the HLA-DR genes but however recently two non-shared epitope alleles DRB*13 and DRB*15 have also been linked to rheumatoid arthritis [131-133]. The involvement of HLA-DR alleles and the identification of MHC class II expressing APCs and T cells show that MHC class II dependent B and T cell activation are the main driving force for the pathological symptoms of rheumatoid arthritis [134,135]. Furthermore, certain non-HLA gene single nucleotide polymorphisms have been linked to rheumatoid arthritis as well. A tyrosine phosphatase, PTPN22, also playing a role in T and B cell signalling, is in fact the second most high genetic risk variant after HLA-DR alleles, linked to rheumatoid arthritis susceptibility [136-138]. Various other genome wide analysis studies from different populations around the world revealed several single nucleotide polymorphisms in non HLA-DR gene loci which are identified as high genetic risk variants for rheumatoid arthritis. These are TRAF1 (C5 locus)- a gene encoding an intracellular protein mediating signal transduction via tumour necrosis factor (TNF) receptors [139,140]; STAT4- a gene encoding a transcription factor which is induced in the presence of cytokines such as IL 12 and IL 23 [141]; OLIG3-AIP3- a gene encoding a crucial factor required for the termination of TNF induced signals [142] and PADI4- a gene encoding calcium dependent enzymes helping in citrullination of peptides [143].

The shared epitopes on HLA-DR alleles and single nucleotide polymorphisms in other gene loci such as PTPN22, TRAF1 (C5 locus), STAT4, OLIG3-AIP3 and PADI4 are considered high genetic risk variants associated with rheumatoid arthritis susceptibility. Further large scale genome wide analysis performed have revealed at least 30 gene loci majority of which are involved in B and T cell signalling and activation are involved in rheumatoid arthritis susceptibility [144].

Environmental factors

Various studies on twins pairs, show that even though the twin off-springs share the same genetic makeup, if one of the twin is exposed to some environmental factors then the risk of developing rheumatoid arthritis is much higher than in the other twin

offspring suggesting highly that environmental and chance factors along with genetic factors play a major role in the pathogenesis of rheumatoid arthritis [145,146]. One of the best studied environmental factors is cigarette smoking. The risk of developing rheumatoid arthritis was shown to be correlating linearly with the intensity and duration of smoking [147,148]. It was shown that once the lungs encounter smoke, activation of macrophages and/or cells enters apoptosis [149]. Furthermore, this led to high activity of enzymes called peptidylarginine deiminases which causes citrullination of various proteins in the lungs [150]. These post-translationally modified proteins bind specifically to shared epitope binding motifs of the HLA-DR molecules on APCs which further leads to complex immune reactions to citrullinated peptides [151,152]. Other environmental risk factors associated with rheumatoid arthritis are exposure to silica dust in granite factory workers [153,154] and other occupational hazards wherein airway exposures to other agents including mineral oils are linked to rheumatoid arthritis [155]. Moreover, certain environmental factors are shown to confer protection to development of rheumatoid arthritis. For example breastfeeding more than 12 months [156] and postmenopausal hormone replacement are known to confer protection not in all but in many cases [157]. Interestingly, moderate alcohol consumption reduces the risk and severity of experimental arthritis in rodents [158] and has been also proven in a human case study wherein consumption of ~80 mg of alcohol per day reduced the risk of rheumatoid arthritis by 40 to 50% compared to non-drinkers [159].

However, little is known if infections or exposure to certain microorganisms could promote or trigger an autoimmune response in rheumatoid arthritis by antigen or molecular mimicry. In one study it was reported that the heat-shock protein dnaJ of *E. coli* possessed similar amino acid sequence (QKRAA) as is present on certain HLA molecules conferring susceptibility to rheumatoid arthritis [160]. Immune responses to the bacterial antigen had stronger affinity in synovial fluid cells of rheumatoid arthritis patients suggesting that auto-reactive T cells cross reacts with autologous heat shock proteins which are highly expressed at synovial sites of inflammation [160]. Mimivirus and other giant viruses infecting the ubiquitously present amoeba have the potential of being a mimicking viral agent in rheumatoid arthritis due to the presence of several collagen-like proteins. Moreover, since it encodes its own glycosylation machinery it can

post-translationally modify its own and potentially but less likely host proteins, increasing the wide range of candidate proteins and/or peptides for molecular mimicry.

Pathogenesis, immune regulations and progression of rheumatoid arthritis

Occurrence of antibodies against collagen type II (CII) and CII-specific T cells clearly suggests that CII is one of the major auto-antigen in rheumatoid arthritis [161,162]. This is also correlating with the fact that CII is the major protein present exclusively in joint tissues and cartilage [67]. CII-specific immunity thus plays a major role in the beginning of inflammation in joints [162,163]. More than just auto-antibodies against CII, as rheumatoid arthritis develops, there are complex immune responses developed within the joint tissue which leads to joint inflammation and destruction (Fig. 16).

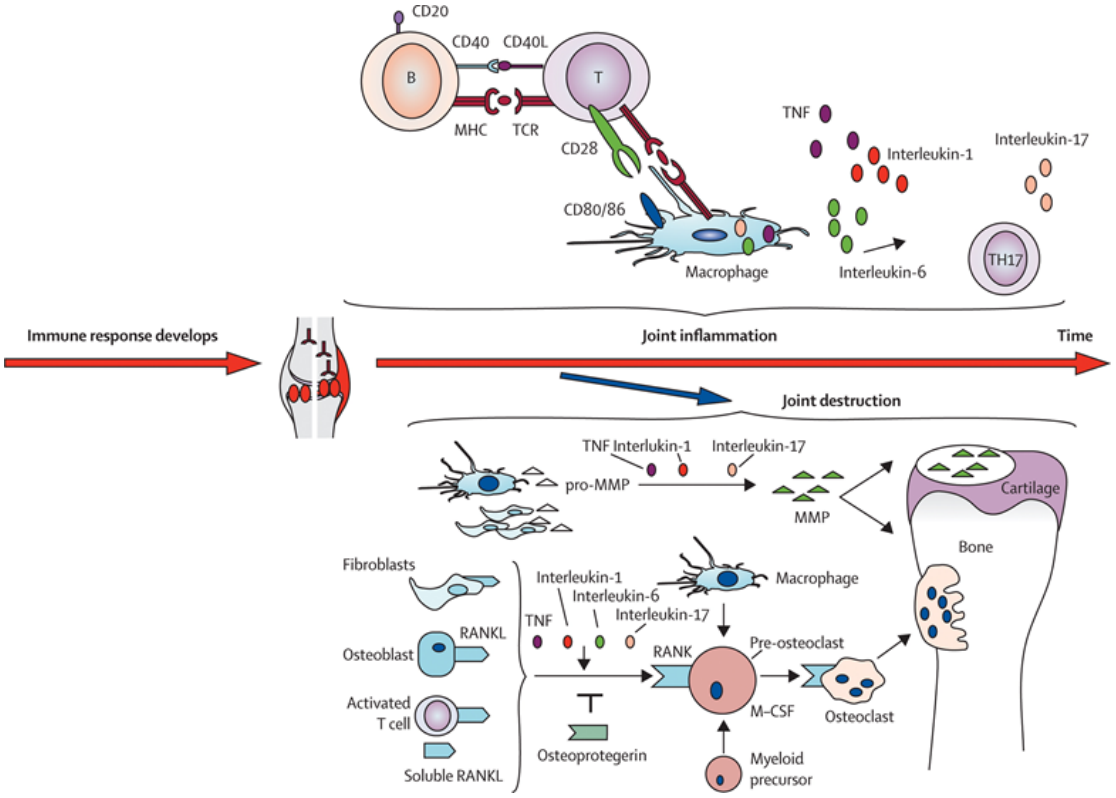


Fig. 16-Phases of joint inflammation and destruction in an arthritic joint. (Figure taken from [127]) Upper part elucidates the interaction between B and T cells with activated macrophages secreting pro-inflammatory cytokines such as TNF, IL 1, IL 6 and IL 17 which leads to joint inflammation. In parallel to joint inflammation, the lower part of the figure elucidates complex immune reactions in the influence of pro-inflammatory cytokines, bone destructing proteases (MMPs) and molecules (RANKL) are over expressed leading to generation of osteoclasts ultimately resulting in bone and cartilage destruction.

One of the major inflammatory responses prompts to the production of cytokines which play an important role in the pathogenesis of rheumatoid arthritis further leading to joint destruction. A wide range of immune cells interact with each other in the inflamed joint synovium producing various pro-inflammatory cytokines. Through the TCR-MHC interaction, APCs communicate with T cells and activate them [127,134]. Only in the presence of co-stimulatory signals mediated via CD28-B7 receptor family (CD80/86) can the T cells be activated [101,134]. B cells are bifunctional- acting as APCs and also as antibody producing cells transporting antibodies to the immune complex formation [127,164]. Pro-inflammatory cytokines such as TNF, IL 6 and IL 1 are secreted by the T cell mediated macrophage activation [164-167]. In fact, analysis of arthritic joint synovia showed the presence of activated T and B cells in majority of inflamed tissue and the presence of TNF, IL 6 and IL 1 [164-167]. On further activation by IL 6, another subset of T cells called T-helper 17 (Th17) cells produces IL 17 which helps in the secretion of cartilage and bone destruction-related enzymes and molecules such as receptor activator of NF-κB ligand (RANKL) [168,169]. The general tiredness or fatigue observed in rheumatoid arthritis patients is because of the action of IL 1 and IL 6 on receptors in the brain cells affecting central parts of the brain [168,169].

Matrix metalloproteinases (MMPs) are released from fibroblasts and macrophages due to the joined effort of pro-inflammatory cytokines TNF, IL 1 and IL 6 [170-172]. Of the 19 MMPs known so far, MMP1 and MMP3 play an important role in destruction of extracellular matrix cartilage [170-172]. Bone erosion and joint space narrowing is one of the main characteristic of rheumatoid arthritis development. Again, pro-inflammatory cytokines TNF, IL 1 and IL 6 drive the expression of RANKL and its further release from fibroblasts, activated T cells and osteoblasts [170-172]. One of the major bone erosion molecules is RANKL which when expressed in large amounts leads to the activation of osteoclasts (from osteoclasts precursors) resulting in bone destruction [170-172].

Treatment strategies towards rheumatoid arthritis

Rheumatoid arthritis being a complex disease, few drug treatment strategies are available targeting specific molecules during the development of disease to reduce the pain and tissue soreness in patients. Some clinical studies in the past have demonstrated that early and aggressive treatment with drugs which are collectively termed as disease-modifying anti-rheumatic drugs such as methotrexate, sulfasalazine, hydroxychloroquine, leflunomide and glucocorticoids which control the inflammatory activity and bone erosions have been proved highly beneficial in rheumatoid arthritis patients [173-176]. Considering the major role played by cytokines in the development of rheumatoid arthritis as mentioned above, biotechnological advances have targeted many of these cytokines and its respective receptors as potential drug targets. Monoclonal antibodies mediating TNF-blocking, neutralize circulating and synovial TNF, have been widely approved and used in clinical trials such as infliximab (chimeric anti-TNF), etanercept (soluble TNF receptor) and adlimumab (humanised anti-TNF) [127,177,178]. Combined therapies of one or two disease-modifying anti-rheumatic drugs and TNF-blocking antibodies have been proved highly beneficial in patients by reducing inflammation and virtually eliminating joint destruction [178-181]. Targeting TNF and its success, led to the discovery of additional biological drugs targeting other pro-inflammatory cytokines. Human IL 1 receptor antagonist, anakinra and IL 6 receptor antagonist, tocilizumab helps competitively inhibiting IL 1 and IL 6 respectively to bind to its receptors and thus leading to reduced bone erosion in patients [182,183]. Many monoclonal antibodies targeting B and T cell interactions have been also approved and widely used. Abatacept, a fusion protein, has proved beneficial in inhibiting co-stimulatory signals required for T cell activation [183-185]. Rituximab binds to CD20 on B cells and helps in reducing them from circulation [183-185]. It is important to note that, different therapies would work differently for individual patients at various time-points [127]. To attain utmost disease control, it is imperative to identify the mode of therapy i.e. monotherapy or combination therapy might be useful for individual patients [127]. Even though various drugs and molecules as mentioned here are available, it is essential to note that, they only reduce inflammation and pain in patients since the treatments are purely symptomatic.

Animal models of rheumatoid arthritis

To study the progression of rheumatoid arthritis and recognize the potential of new drug targets towards rheumatoid arthritis, it is important to identify an ideal *in vivo* animal model which could closely resemble the human rheumatoid arthritis which as explained above is polygenic and depends on various environmental factors [186]. Animal models are useful as both genetics and environment can be better controlled [186]. Most of the animal models of rheumatoid arthritis are established in rodents (rats or mice) and are broadly classified into two types: inducible models and genetically modified spontaneous models.

Collagen induced arthritis

One classical inducible mouse model closely resembling rheumatoid arthritis in humans and widely accepted as the conventional *in vivo* model is collagen induced arthritis (CIA) wherein B and T cell recognition to CII can be elicited as in human rheumatoid arthritis [187,188]. Considering the exclusivity of CII expression in articular cartilage joints, onset of arthritis can be observed in mice when immunized with heterologous species CII, emulsified in an adjuvant generally complete Freund's adjuvant (CFA), inducing joint inflammation and destruction [187-189]. Susceptibility to rheumatoid arthritis, in humans, is associated with HLA alleles DRB1*01 and DRB1*04 [131-133]. Mice expressing similar haplotypes I-A^q and I-A^r such as DBA/1 and B10.Q have been shown to be highly susceptible to CIA [190,191]. The peptide binding pocket of A^q molecule is highly similar to the shared epitope of DRB1*01 and DRB1*04, conferring it susceptible to CIA [190-193]. Furthermore, the immunodominant peptide of CII molecule between positions 260-270 is shown to bind with the A^q molecule [194]. This glycopeptide is readily recognized by TCRs of CII specific T cells [85,195]. Along with this genetic susceptibility in mice, anti-CII specific IgG antibody levels are also high and accumulation of these antibodies in the synovium further leads to complex inflammation responses [196,197]. Furthermore, pro-inflammatory cytokines such as TNF, IL 6 and IL 17 are secreted by the infiltrating macrophages and activated T cells [198]. IL 17 in turn has been shown to activate many matrix degrading enzymes such as MMPs, cathepsin G and other matrix degrading molecules such as RANKL which help in

the production of osteoclasts resulting in destroying the collagen matrix and bone erosions [199,200].

The genetic susceptibility in mice strains coupled with engagement of the immune system activating B cells producing auto-antibodies towards CII and activating T cells producing various pro-inflammatory cytokines and matrix-degrading enzymes in the synovium of CIA mice, breaking immune tolerance, resulting in joint inflammation and destruction as observed in humans, proves to be an ideal model for rheumatoid arthritis [196-200].

Collagen-antibody-induced arthritis

As seen above, rheumatoid arthritis is associated with an auto-antibody production towards CII, it was observed if non-immunized mice are immunized with the serum of CIA mice (wherein auto-antibodies towards CII are detectable), arthritis can be induced in the recipient non-immunized mice [201]. Furthermore, this antibody transfer suggests clear role of humoral immunity in the development of arthritis [202]. Anti-collagen antibody cocktails are now available to induce arthritis in mice allowing better understanding of humoral auto-immunity aspect of rheumatoid arthritis [202]. Even though the clinical development of arthritis in mice by collagen-antibody-induced arthritis is similar to rheumatoid arthritis and CIA, it is not accompanied with B and T cell response but rather depends on polymorphonuclear inflammatory cell infiltration [202]. However, it was shown that disease severity was enhanced if anti-CII reactive T cells were administered to the mice [203]. In totality, using this model the distinct roles of innate and adaptive immune system towards development of arthritis can be studied clearly [202].

Zymosan-induced arthritis

A polysaccharide from the yeast *Saccharomyces cerevisiae* called zymosan (repeating glucose units in β-1,3-glycosidic linkage) has been also used to induce arthritis in mice [204]. Zymosan is known to activate macrophages and secrete pro-inflammatory cytokines and activate the complement cascade via an alternate pathway and thus induce arthritis in mice [204]. The injection of zymosan in the knee joints of mice results in a chronic quick development of arthritis (generally between 3-5 days after

injection) and is associated with proliferative inflammatory arthritis and synovial hypertrophy [204]. The major limitation of this model is the technical expertise required to perform an intra-articular injection and furthermore the injection directly in the joint articular tissue prevents the study of the systemic part of the disease [202].

Streptococcal cell wall arthritis

The peptidoglycan polysaccharide component of streptococcal cell wall when injected intra-peritoneally results in chronic, erosive polyarthritis in Lewis rats [205]. In comparison to Lewis rats, the same model in mice is less robust and chronic and is known to only show the acute phase of the disease [206]. However, the chronicity and severity of inflammation could be enhanced with intra-articular injections rather than intra-peritoneal injections [207]. This model was shown to be associated with T cell response with infiltration of neutrophils, activated macrophages and pro-inflammatory cytokines TNF, IL 1, IL 6 [205,207].

Genetically modified spontaneous models

Various mouse strains are established to develop spontaneous arthritis without inducing it with any specific antigen or adjuvant. Initial work showed that over expressing human TNFα in mice led to the development of chronic inflammatory polyarthritis [208]. Along with CIA, this model also closely resembles rheumatoid arthritis in humans showing chronic progressive disease. Treatment with a monoclonal antibody targeted towards TNF completely prevented the onset of the disease [208]. This model also helped researchers to build up therapies against rheumatoid arthritis in humans targeting TNF as mentioned above. Transgenic mouse lines such as K/BxN (generated by crossing TCR transgenic mice with non-obese diabetic mice) are shown to develop spontaneous severe destructive arthritis [209]. The mechanism involved in the development of arthritis is associated with complement activation and mast cell degranulation which is regulated by TNF and IL 1 [209]. Since this mouse model shows high titers of antibodies recognizing glucose-6-phosphate isomerase, due to the recognition by transgenic TCR, antibodies in human rheumatoid arthritis patients recognizing the same isomerase and linking it to pathogenesis of the disease was quite debatable and controversial [210]. Other transgenic mouse models developing

spontaneous inflammatory arthritis include mice deficient in IL 1 receptor antagonist which is mediated by Th17 cell response [211] and implantation of human rheumatoid arthritis synovial tissue grafts in severe combined immune-deficient mice helping to analyse the destructive behaviour of synovial fibroblasts in an animal model [212].

In whole, these different mouse models gives us insights into the many mechanisms of arthritis in animals but clearly shows the similarities between models in the inflammatory processes involved in the pathogenesis of rheumatoid arthritis.

Aims and objectives

CII is the major protein present exclusively in the joint articular tissues and is the major auto-antigen identified in rheumatoid arthritis [67,161,162]. Anti-CII specific B and T cells are recognized in rheumatoid arthritis synovial tissues and thus CII specific immunity plays a major role in joint inflammation and destruction [162,163]. Mimivirus possesses seven ORFs encoding collagen-like proteins [3], of which at least one L71 is expressed at the surface of mimivirus particles (Shah *et. al.*, 2013). Furthermore, an immunodominant peptide sequence in CII recognized by T cells in rheumatoid arthritis [85,86,195] has sequence homology in L71 as well (Shah *et. al.*, 2013). Giant viruses like mimivirus and others have been routinely isolated from natural environments from various ecosystems elucidating that such viruses are present in the environment and humans are at a potential risk of exposure to them [1,54,55]. Thus, considering the environmental risk of coming in contact with giant viruses, and possibly by the phenomenon of antigen or molecular mimicry between human and mimivirus collagens, we hypothesized the potential of mimivirus triggering an autoimmune response like rheumatoid arthritis. Under this the specific aims for the project were:

- To characterize the immune response *in vivo* in CIA mouse model to elucidate the potential of mimivirus in eliciting an autoimmune response like arthritis in mice

- To determine exposure to mimivirus, human sera samples were analysed for anti-mimivirus antibodies

- To identify specific viral proteins recognized by human sera so as to confirm the specificity of antibodies to mimivirus proteins

Materials and Methods

(The majority of the methods presented here are part of the manuscript Shah *et. al.*, 2013)

Ethics statement

All mouse experiments were performed in compliance with the Swiss Animal Protection Ordinance and approved by the local veterinary authority (Kantonales Veterinäramt Zürich, Switzerland). The human sera tested in this study were provided by Michel Neidhart (Center of Experimental Rheumatology, University Hospital Zurich), were a part of previously existing collection and the experimental protocol approved by the Kantonale Ethik-Kommission Zürich (KEK).

Mimivirus infection

Acanthamoeba polyphaga and mimivirus cultures were provided by Didier Raoult (Aix-Marseille Université). Amoeba were routinely cultured as monolayer in PYG (peptone, yeast, glucose) medium at 28°C as previously described [3]. Mimivirus was added to multiplicity MOI 10 to amoeba and newly formed virus was collected from the culture supernatant 2 days post infection.

Mimivirus protein preparation

Mimivirus particles were suspended in 0.5 M Tris-HCl, pH 8.5, 0.2% CHAPS, 2 mM TCEP, 6 M guanidine hydrochloride and incubated at 65°C for 10 min. After cooling it to room temperature, iodoacetamide was added to a final concentration of 3 mM and further incubated at room temperature in dark for 40 min. After adding DTT to a final concentration of 15 mM, protein extracts were centrifuged at room temperature at 17,000xg and proteins in the supernatant were precipitated with 12% trichloroacetic acid.

CIA mouse model

DBA/1 mice were purchased from Charles River (Germany), bred and maintained in the animal facility of Institute of Physiology, University of Zurich under standard conditions.

CIA was established as described before [189]. Briefly, 6-8 week old mice were immunized intra-dermally in the tail either with PBS or bovine collagen type II (Chondrex, USA) or UV-inactivated mimivirus protein extract emulsified in CFA (Chondrex, USA). Each mouse received 50 µl PBS or 100-120 µg of bovine collagen type II or 120-150 µg of mimivirus proteins emulsified 1:1 in CFA in a total volume of 50-70 µl. Mice received 30 days later a booster injection of the same amount of antigen emulsified 1:1 in Incomplete Freund's Adjuvant (IFA, Chondrex, USA). Development of arthritis was monitored daily for at least 75 days post immunization. Severity was scored on a level of 0 (no inflammation) to 4 (most severe inflammation) per limb per mouse, thus allowing a maximum score of 16 per mouse [189].

Anti-CII antibodies

Mouse blood sera were collected by heart puncture. Anti-mouse CII and anti-human CII antibody titers were measured in blood sera, by an ELISA kit (Chondrex, USA) and performed as per manufacturer's instructions.

Histology

Limbs were prepared by removing skin and fixed overnight in 10% neutral buffered formalin. Tissues were further decalcified in Immunocal solution (Quartett, Germany) for 4-5 days, dehydrated and paraffin embedded. Sections of 5 µm were mounted on glass slides, stained with H&E and analysed with light microscopy.

Recall assay

Draining lymph nodes (axillary, lateral axillary, superficial inguinal and popliteal) from mice were collected 8-10 days post booster injection. 100,000 cells in 100 µl complete RPMI-1640 medium were stimulated with antigens and incubated in a CO_2 incubator at 37°C with 5% CO_2. Cells were stimulated in 100 µl either with medium alone (negative control) or concanavalin A (Sigma, Switzerland) at 3 µg/ml (positive control). T-cell proliferation grade denatured mouse collagen type II at 1 mg/ml (Chondrex, USA) or T-cell proliferation grade denatured bovine collagen type II at 1 mg/ml (Chondrex, USA) and 45 min 95°C treated denatured mimivirus collagen L71 at 1.5 mg/ml were used as

antigens. After 48 h, 1 µCi of [^{3}H] thymidine (Perkin-Elmer, USA) per well was added, incubated for 16-18 h and cells were harvested on 96-well glass filter (Perkin-Elmer, USA) and assayed for incorporation of [^{3}H] thymidine using a 96-well scintillation beta counter (Wallac, Perkin-Elmer). All tests were done in duplicates.

In vivo imaging

One day prior to scarifying the CIA mice, they were injected intravenously in the tail vein with 4 nmol per mouse of ProSense 750 (Perkin-Elmer, USA) florescent imaging agent. 24 h later, the mice were imaged for inflammation using the IVIS spectrum available at the Zurich Integrative Rodent Physiology (ZIRP), Institute of Physiology, University of Zurich.

Anti-mimivirus ELISA

Mimivirus protein extract were coated in microtiter plates at 0.1 µg per well in 100 µl PBS overnight at 4°C. Plates were washed thrice with PBS-0.05% Tween and blocked with PBS-0.05% Tween, 1% bovine serum albumin, 0.05% NaN_3 at 37°C for 2 h. Plates were washed thrice, 100 µl of human or rabbit sera at various dilutions were added and further incubated at room temperature for 1 h. All dilutions were performed in PBS-0.05% Tween, 0.1% bovine serum albumin. Plates were washed thrice, 100 µl of 1:5000-diluted biotinylated anti-human or anti-rabbit IgG or IgM antibody (BD Biosciences, Switzerland) were added and incubated at room temperature for 2 h. Plates were washed thrice and 100 µl of 1:1000-diluted streptavidin-HRP conjugate (BD Biosciences) were added and incubated further at room temperature for 1 h in dark. Plates were washed thrice, incubated for 2 min in dark with 50 µl 3, 3', 5, 5' tetramethylbenzidine (TMB) substrate (BD Biosciences) before stopping the reaction with 25 µl of 2 N H_2SO_4. Plates were analysed at 440 nm. All sera were tested in duplicates.

Immunoprecipitation of mimivirus proteins

Aliquots of 25 µl of human sera were incubated with 30 µl of protein-G sepharose 4 Fast Flow (GE Healthcare, Switzerland) beads along with 80 µl 20 mM sodium phosphate, pH

7 on a rotating shaker for 1 h at 4°C. After centrifugation at 500 x g for 5 min at 4°C, supernatants were discarded and beads incubated with 20 µg of mimivirus protein extract in 80 µl 20 mM sodium phosphate, pH 7 and further incubated on a rotating shaker for 30 min at 4°C. Beads were washed thrice with 20 mM sodium phosphate, pH 7 and antigen-antibody complexes were eluted from the beads by adding 40 µl of 0.1 M glycine, pH 2.7. After neutralization by addition of 20 µl of 1 M Tris-HCl pH 9, eluates were separated by SDS-PAGE and stained overnight in Rapid stain solution (G Biosciences, USA). Slices of polyacrylamide gel excluding IgG chains were excised and subjected to in-gel tryptic digest as previously described [213] and peptides identified by tandem mass spectrometry and validated by Mascot software (Matrix Science, London, UK; version 2.3.02). Mascot was set up to search a Swissprot concatenated target-decoy database (2011.01.11, 1049100 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 PPM. Iodoacetamide derivative of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine was specified in Mascot as variable modifications.

Cloning, bacterial expression and purification of recombinant proteins

Mimivirus genomic DNA was isolated from purified virus particles by TRIZOL reagent (Invitrogen) and performed as per manufacturer's instructions. Mimivirus collagen-like protein L71 was cloned from genomic DNA using the primer pair 5'-TGACCTCGAGTCAAGAATAACTTGCCCAATTAC-3' and 3'-CAGCAAGCTTATTTATGCAAGAGCTAATCCAG-5'. The resulting 2843 bp fragment was first cloned into pBluescriptsK+ and further sub-cloned into the expression vector pET16b (Merck) using the *Xho*I and *Hind*III sites. The recombinant plasmid was transformed into BL21 (DE3) *E. coli* cells (Novagen, Switzerland) and selected using ampicillin (100 µg/ml) supplemented in LB-agar plates. Mimivirus major capsid protein L425 was cloned from pFastBac vector which was earlier cloned in our lab, using the primer pair 5'-CTCGAGATGCTTGGTGACGTACCTGA-3' and 3'-GGATCCTCAATTACTGTACGCTAATCCG-5'. The resulting 1368 bp fragment was first cloned into pJET vector (Fermentas) and further sub-cloned into the expression vector pET16b (Merck) using the *Xho*I and *BamH*I sites. The recombinant plasmid was

transformed into BL21 (DE3) pLysS *E. coli* cells (Novagen) and selected using ampicillin (100 µg/ml) supplemented LB-agar plates. A fragment of human GLT25D2 galactosyltransferase [73] was truncated using *BsrG*I and used as negative control in Western blot analysis.

Cells transformed with the plasmid were expressed in 300-500 ml cultures and induced with 0.2 mM IPTG at 32°C for 1.5 h. The recombinant His_6-tagged proteins from the soluble fractions were purified over Ni-Sepharose 6 Fast Flow (GE Healthcare) beads by gravity flow column purifications. After washing the bound protein to the beads with 10 column volumes with wash buffer (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4), proteins were eluted using 3 column volumes of the same buffer containing 200 mM imidazole for L71, 500 mM imidazole for L425 and 100 mM imidazole for truncated GLT25D2 protein.

Western blotting

Aliquots of 15 µg of recombinant mimivirus L71 and L425 proteins were subjected to SDS-PAGE and transferred to PVDF membrane (Bio-Rad). A His_6-tagged 15 kDa fragment of the truncated human GLT25D2 protein was used as negative control. A 36 kDa His_6-tagged fragment of human collagen type III encompassing 114 [G-X-Y] repeats and lacking N- and C-propeptides was provided by Christoph Rutschmann (Institute of Physiology, University of Zurich). Blots were blocked in 1% polyvinylpyrrolidone (Sigma, Switzerland) + 5% dry milk solution overnight at 4°C, then washed three times for 5 min with TBS, 0.1% Tween and incubated with human sera diluted 1:4000 for 2 h at room temperature. After washing four times for 5 min, blots were incubated with anti-human IgG-HRP (Promega, Switzerland) at 1:7500 dilution at room temperature for 1 h. Blots were developed with SuperSignal chemiluminescent substrate (Thermo Scientific).

Statistical analysis

One way ANOVA with Dunnett's multiple comparison (GraphPad Prism) was performed to compare experimental groups.

Isolation of giant viruses from environment

One litre of water sample from 14 sources around Zurich were collected (Table 6). Each water sample was first filtered through 0.8 µm filter (Millipore) and the filtrate was further filtered through 0.2 µm filter (Millipore) to capture any giant viruses present in the water samples. The 0.2 µm filter was cut from the filter unit and incubated with 2 ml PYG medium with rotation for 3-4 days at 4°C.

Table 6-Collection points for water sample analysis

Sample #	Collected from	Sample #	Collected from
1	Pond in forest, Opfikon	8	Greifensee
2	Biotope water, Opfikon	9	Furtbach, Otelfingen
3	Tessin lake	10	Artificial lake, Opfikon
4	Katzensee	11	Fountain water, Otelfingen
5	Zurich lake, Bellevue	12	Limmat
6	Fountain water, Rentenwiese	13	Irchel pond
7	Pond E floor, Irchel campus	14	Zurich lake, Rentenwiese

The culture supernatant was used to infect flasks of amoeba *A. polyphaga* which were supplemented with appropriate antibiotics mix of penicillin (5 units/µl) and streptomycin (5 µg/µl) only (for samples 1-7) or a mix (for samples 8-14) of penicillin (5 units/µl), streptomycin (5 µg/µl), ampicillin (100 µg/ml), kanamycin (50 µg/ml) and gentamycin (70 µg/ml) to dampen the growth of bacteria. After 4 days of incubation, the supernatant was collected and genomic DNA was extracted by standard ethanol precipitation method. This DNA was used as a template to perform PCR analyses with mimivirus degenerative primer targeting polymerase B1 (Pol B1) with primer pair 5'-

CCATTMAARATTATGGGTTATGATATTG-3’ and 5’-ATATTTGTCGTCAAAACCAAAATTATT-3’ and degenerative primer pair targeting polymerase B2 (Pol B2) with primer pair 5’-TWGATGTHACAGAAGATCATTC-3’ and 5’-CAGTTTCCARATCATAYACATAWTCTC-3’ and marseillevirus (T19 strain) degenerative primer targeting AVS polymerase (AVSpol) with primer pair 5’-GARGGNGCNACNGTNYTNGAYGC-3’ and 5’-GCNGCRTANCKYTTYTTNSWRTA-3’. The extracted DNA of positive samples was also used for whole genome sequencing which was performed using the on the *de novo* sequencing apparatus (Pacific Biosciences RSII) available at the Functional Genomics Centre Zurich. The sequences retrieved were compared and aligned on the Unipro UGENE software (version 1.11.5). The virus isolates from culture supernatants of infected amoeba were harvested as described above and epoxy resin embedded ultrathin sections were observed by using TEM Philips CM208 electron microscope (Centre for Microscopy and Image Analysis, University of Zurich) operating at 100 kV.

Results

(The majority of the results presented here are part of the manuscript Shah *et. al.*, 2013)

The broad occurrence of giant viruses in the environment and structural similarities with mammalian proteins led us to analyse the possibility of mimivirus in initiating an autoimmune response. Firstly, in the CIA mouse model, mice were immunized with mimivirus protein extract to observe whether mimivirus has the potential to initiate an arthritis-like response in mice. Furthermore, to understand the perspective of humans being exposed to giant viruses, human sera were analysed for antibodies against mimivirus proteins and specific viral proteins being recognized were identified. Additionally, we tried to isolate giant viruses from local natural aquatic environments and this gives an idea that viral members of *Mimiviridae* and *Marseilleviridae* families of NCLDV group are present in them.

Mimivirus proteins promotes arthritis in mice

Joint inflammation in mice immunized with mimivirus proteins

Mimivirus possesses seven ORFs representing collagen-like proteins and we also identified that at least one of the collagen-like proteins L71 is expressed at the surface of mimivirus particles (Shah *et. al.*, 2013). For this reason, first we studied the potential of mimivirus to promote joint inflammation in mice in the CIA mouse model which closely resembles rheumatoid arthritis in humans. For this study DBA/1 mice were used since it was shown to possess similar susceptible HLA haplotypes (I-A^q and I-A^r) as in humans which are linked to rheumatoid arthritis rendering these mice susceptible to CIA [190,191]. As a positive control mice were immunized intra-dermally with bovine CII and as a negative control mice were immunized with PBS emulsified in Freund's adjuvant. Immunization of mice intra-dermally with bovine CII and mimivirus proteins led to joint inflammation as observed by anonymous visual clinical scoring of mice limbs and histological examination. Clinical severity scores by 75 days post immunization reached up to 6 for mice immunized with mimivirus proteins and up to 12 for mice immunized with bovine CII (Fig. 17). As expected, the negative control mice which were immunized with PBS did not show any clinical severity. Even though the severity of arthritis in mice immunized with mimivirus protein extracts was not as high as our

positive control, this observation was interesting since mice did show phenotypic manifestation of the autoimmune disease after immunization with mimivirus proteins.

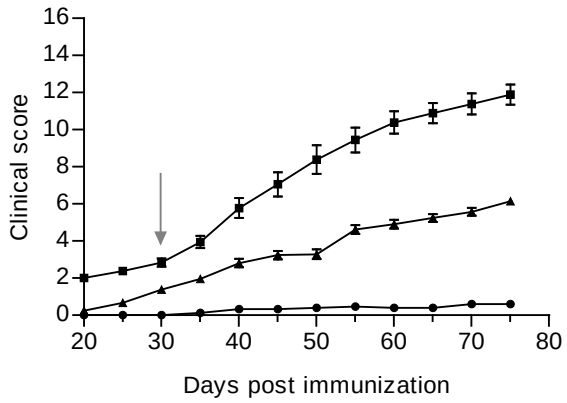


Fig. 17-Joint inflammation in DBA/1 mice immunized with mimivirus proteins. Mice were immunized with either PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). The clinical severity of arthritic limbs in the groups of Neg (circles), BovCII (squares) and MV (triangles) immunized mice are shown as mean ± SEM. The arrow shows the time point of booster immunization. Data represent three independent experiments including 15-21 mice per group.

Furthermore, an altered cartilage integrity and synovial hyperplasia in the joints was found in mice immunized with bovine CII and also mice immunized with mimivirus proteins but to a lesser extent (Fig. 18). Infiltration of cells was also observed in these two groups of immunized mice and it is tempting to speculate that these cells were infiltrating leukocytes or macrophages or osteoclasts precursor cells but however this cannot be confirmed by H&E staining alone.

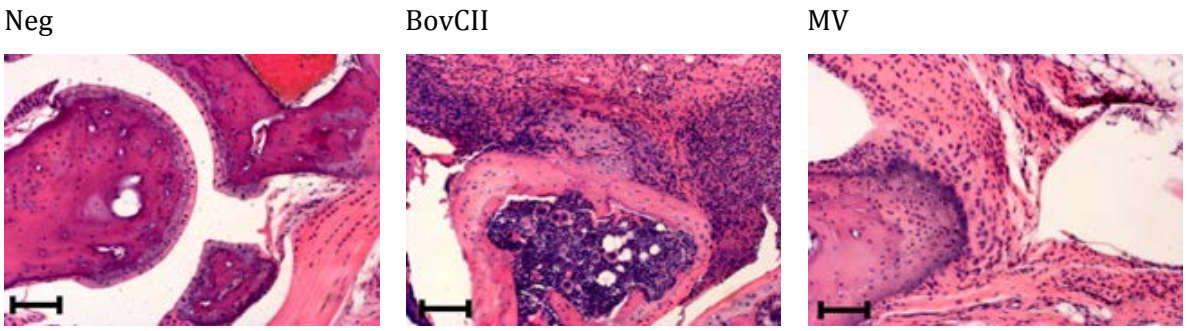
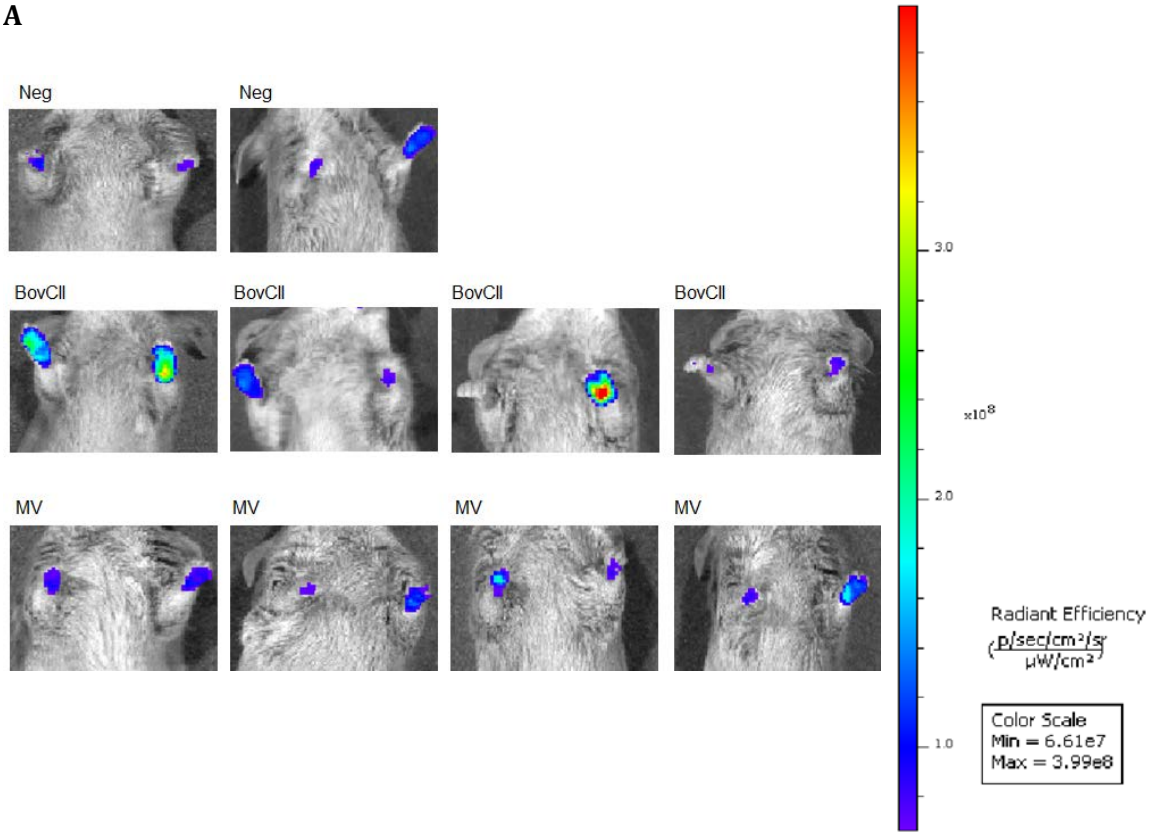


Fig. 18-Histological examination of joint inflammation in immunized DBA/1 mice. Mice were immunized with either PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). Representative H&E stained sections of hind limbs by day 75 after immunization showing cartilage damage and synovial hyperplasia in BovCII and MV immunized mice. No sign of pathology were visible in negative control mice (Neg), scale bar 100 μm.

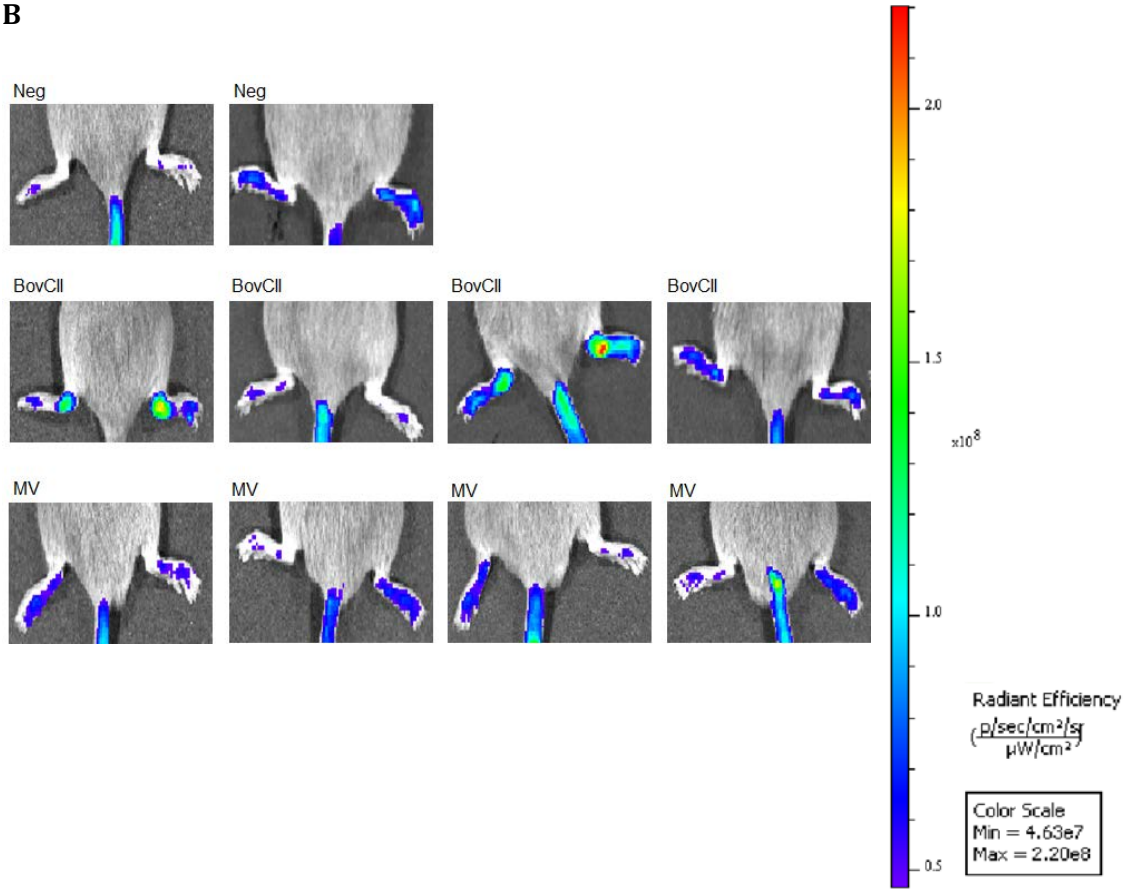
Localization of joint inflammation in joint articular tissues

To check for the localization of inflammation, *in vivo* fluorescence imaging technique was used wherein the fluorescent dye used is highly target specific. Only when the dye comes in contact with inflammation associated proteases especially cathepsins, the dye gets activated and thus fluorescence's which can be monitored and quantified. Cathepsins are produced in joint articular tissues during inflammation in arthritis in mice [199,200]. The bovine CII and mimivirus proteins immunized mice showed targeted fluorescence in the fore limbs (Fig. 19A) and hind limbs (Fig. 19B) however, less fluorescence was observed in limbs of PBS immunized mice. The tails of mice also showed inflammation but this is normal since there is always a local inflammation at the site of immunization and thus this fluorescence was not considered during quantification. The fluorescence signal detected in the limbs of mice can be quantified on the IVIS spectrum (higher the radiant efficiency higher the amounts of inflammation associated proteases at the activated sites). A trend of higher radiant efficiency was observed in mice immunized with bovine CII and mimivirus proteins in comparison with PBS immunized mice (Fig. 19C and D). However, this was not statistically significant mainly due to the low number of animals used in this experiment.

A


Neg
Neg
BovCll
BovCll
BovCll
BovCll
MV
MV
MV
MV
3.0
2.0
1.0
x10^8
Radiant Efficiency
(p/sec/cm²/sr / µW/cm²)
Color Scale
Min = 6.61e7
Max = 3.99e8

B


Neg
Neg
BovCll
BovCll
BovCll
BovCll
MV
MV
MV
MV
2.0
1.5
1.0
0.5
x10^8
Radiant Efficiency
(p/sec/cm²/sr / µW/cm²)
Color Scale
Min = 4.63e7
Max = 2.20e8

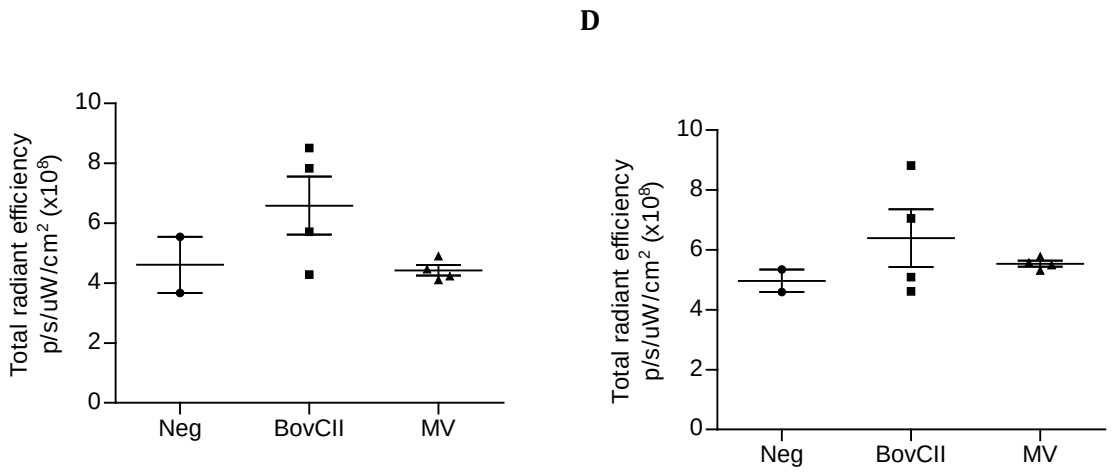


Fig. 19-Inflammation in joints of immunized DBA/1 mice. Mice were immunized with either PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). One day prior to sacrificing the mice, the fluorescent dye was injected intravenously and 24 hours later the mice were imaged live. (**A**) Localization of inflammation as observed by *in vivo* fluorescence imaging in fore limbs of mice. (**B**) Localization of inflammation as observed by *in vivo* fluorescence imaging in hind limbs of mice. The colour scale bar is shown on the right of each figure. The inflammation observed in limbs of immunized mice was marked as region of interest on the quantification software. (**C**) Total radiant efficiency measured in fore limbs of mice. (**D**) Total radiant efficiency measured in hind limbs of mice. n=2 for Neg and n=4 for BovCII and MV immunized mice. Horizontal bars show mean ± SEM.

Breakage of immune tolerance in immunized mice

The characteristic feature of autoimmune diseases is the breakage of immune tolerance towards self-proteins. In arthritis, immune tolerance to endogenous CII is broken and autoantibodies are developed. Since these autoantibodies are targeted against CII, which is exclusively found in joint articular tissues, the joints are the most affected during the disease as seen in the observations above. The breakage of immune tolerance in mice was induced by detecting elevated titers of IgG targeted towards endogenous CII in mimivirus proteins and bovine CII immunized mice (Fig. 20). As expected in the CIA model, the positive control mice immunized with heterologous CII (here bovine CII) showed very high titers of autoantibody IgG response against autologous CII as determined by ELISA. But was surprising to also observe high titers of IgG autoantibodies in mimivirus proteins immunized mice, indicating the immune response

is possibly targeted towards mimivirus collagens. This suggests the development of autoimmunity in mice immunized with mimivirus proteins and bovine CII but not in the negative control mice immunized with PBS. This observation can also be correlated with the clinical scoring. Since there were no autoantibodies to CII detected in the negative control mice, they also did not show any clinical severity. The positive control mice showed extreme autoantibody response to CII and thus also show severe clinical severity. The mimivirus proteins immunized mice displayed an autoantibody response to CII but not as high as the positive control group and is similarly also represented in the clinical scoring.

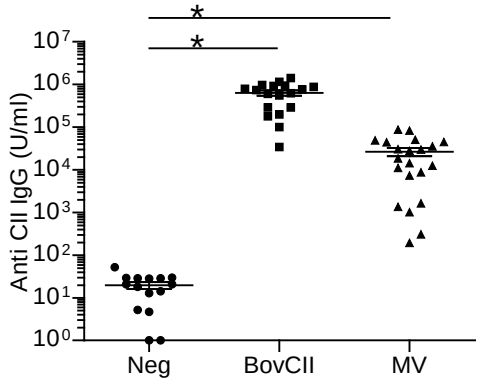


Fig. 20-Anti-collagen type II IgG titers in DBA/1 mice immunized with mimivirus proteins. Levels of serum IgG measured by ELISA against endogenous mouse collagen type II (CII) in mice immunized with PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). Data represent three independent experiments including 15-21 mice per group, horizontal bars show mean ±SEM, *p<0.01.

After analysing the auto-reactive antibody response, auto-reactivity of T cells were tested in recall assays. Cells isolated from the draining lymph nodes of the three groups of immunized mice 10 days after booster immunization, were stimulated *in vitro* with denatured collagen proteins. Cells were seen to proliferate when presented *in vitro* with denatured mouse CII (Fig. 21A), denatured bovine CII (Fig. 21B) and even to denatured mimivirus collagen-like protein L71 (Fig. 21C). This proliferation of cells shows the presence of auto-reactive T cells after immunization of mice. Interestingly, cells isolated from mimivirus proteins immunized mice showed the highest proliferative response to collagen molecules. This observation is surprising because mice immunized with bovine CII showed higher titers of IgG and also showed higher clinical scoring than mimivirus proteins immunized mice. The stronger proliferative response in cells from mimivirus

proteins immunized mice may reflect the differences between the peptide sequences between mimivirus and mammalian collagen sequences.

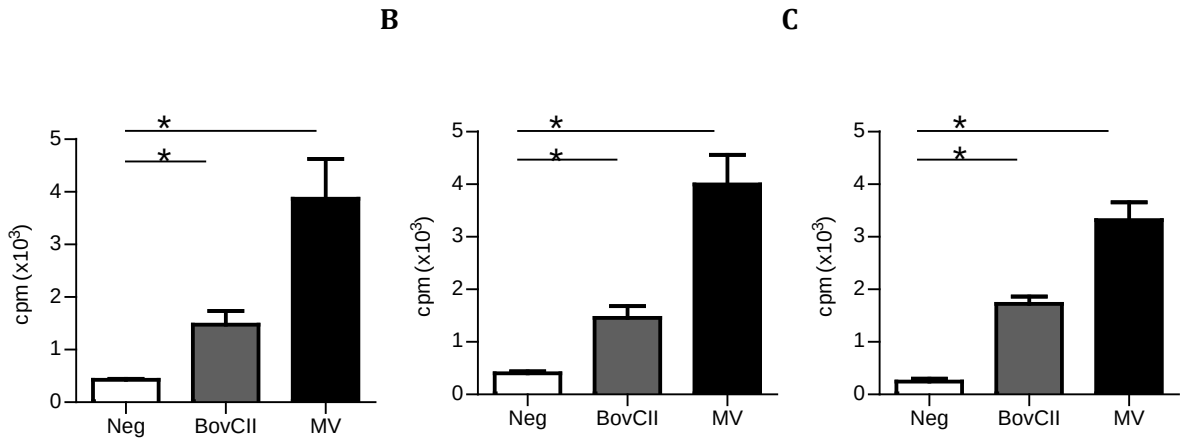


Fig. 21-Auto-reactive T-cell response in DBA/1 mice immunized with mimivirus proteins. (**A**) Recall responses in cells isolated from draining lymph nodes from mice immunized with PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV) after stimulation with denatured mouse collagen type II. (**B**) Recall responses after stimulation with denatured bovine collagen type II. (**C**) Recall responses after stimulation with denatured fragmented recombinant mimivirus protein L71. Data represent mean ± SEM from groups of 3 mice, *p<0.01.

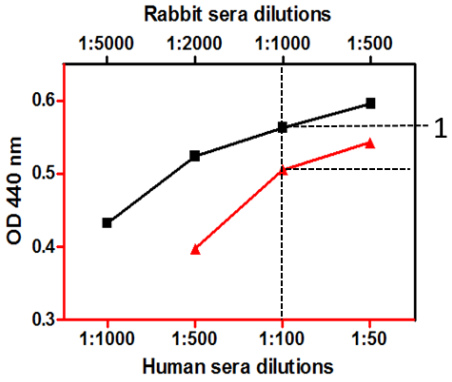
Immunity to mimivirus in humans

Exposure to mimivirus in humans

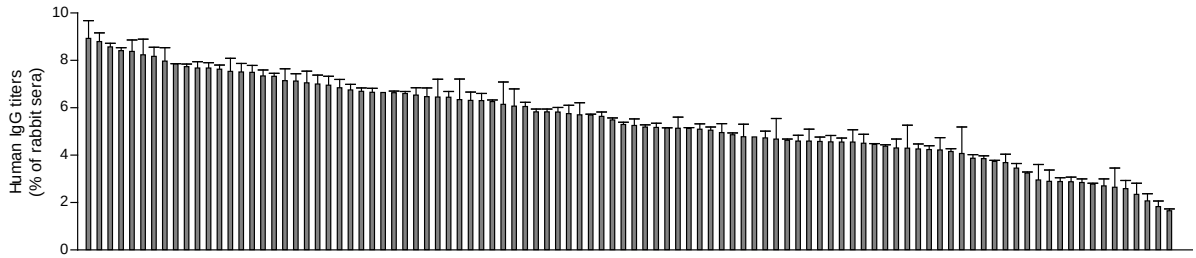
Additionally, to determine whether humans are commonly exposed to mimivirus, the presence of antibody titers against whole mimivirus proteins as antigens were checked in four cohorts of human sera samples by ELISA. The four cohorts included healthy subjects sera, rheumatoid arthritis patient's sera, systemic sclerosis patient's sera (a form of autoimmune disease mainly affecting skin tissue) and osteoarthritis patient's sera (a form of arthritis but not autoimmune in nature). Reactivity to mimivirus proteins within the 100 healthy subject's sera was variable and was observed that 58% of healthy subject's sera showed significant IgG titers in 5% range of titers observed in sera of rabbits previously immunized with mimivirus proteins (Fig. 22A and B). IgG

titers also reached in the 5% range of titers as observed in sera of rabbits in 75% rheumatoid arthritis, 50% systemic sclerosis and 60% osteoarthritis patient's sera as well (Fig. 22C).

A



B



C

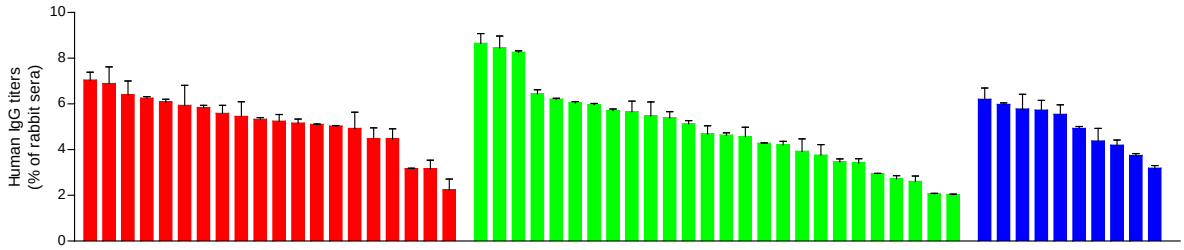
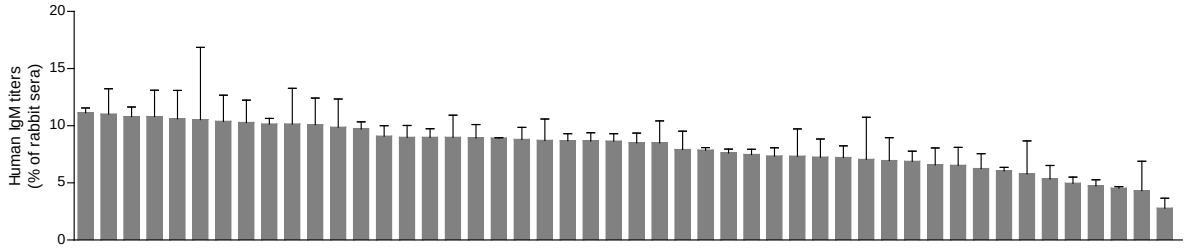


Fig. 22-Reactivity of human sera towards mimivirus proteins. (**A**) Representative plot showing the response of human sera at dilution 1:100 (red line) compared to the response of rabbit sera at dilution 1:1000 (black line). (**B**) Anti-mimivirus IgG titers in sera of healthy subjects (n=100) were measured by ELISA and expressed as a ratio to IgG titers measured in rabbits previously immunized with mimivirus particles. (**C**) Anti-mimivirus IgG titers in sera of rheumatoid arthritis patients (n=20, red bars), systemic sclerosis patients (n=26, green bars) and osteoarthritis patients (n=10, blue bars) measured by ELISA and expressed as a ratio to IgG titers measured in rabbits previously immunized with mimivirus particles. In **B** and **C** error bars show SEM from two independent experiments, each performed in duplicates.

Furthermore, IgM titers were also measured in the various sera samples since it is the most basic antibody produced by the immune system when encountering a foreign antigen. IgM titers reached in the 7% range of titers as observed in sera of rabbits in 75% of a subset of healthy subject's cohort so tested (Fig. 23A). In all rheumatoid arthritis, 30% systemic sclerosis and all osteoarthritis sera IgM titers were highly elevated and reached in the 10% range of titers observed in sera of rabbits (Fig. 23B). These ELISA results show that IgG and IgM titers against mimivirus proteins are present not only in diseased conditions but also in healthy subjects suggesting humans are indeed exposed to mimivirus.

A



B

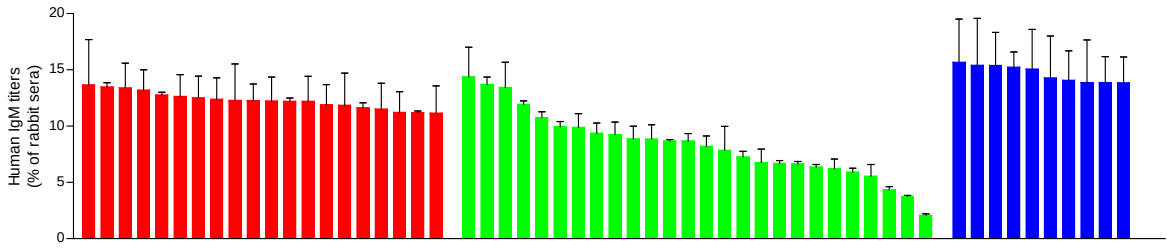


Fig. 23-IgM response of human sera towards mimivirus proteins. (**A**) Anti-mimivirus IgM titers in sera of healthy subjects (n=48) were measured by ELISA and expressed as a ratio to IgM titers measured in rabbits previously immunized with mimivirus particles. (**B**) Anti-mimivirus IgM titers in sera of rheumatoid arthritis patients (n=20, red bars), systemic sclerosis patients (n=26, green bars) and osteoarthritis patients (n=10, blue bars) measured by ELISA and expressed as a ratio to IgM titers measured in rabbits previously immunized with mimivirus particles. Ratios were calculated as shown in Fig. 22A. Error bars show SEM from two independent experiments, each performed in duplicates.

Specific mimivirus proteins recognized by human sera

IgG and IgM antibodies towards whole mimivirus protein extracts were detected in human sera. Now, to further identify the specific major mimivirus proteins recognized by human sera, immunoprecipitation was performed. IgG fractions of sera from healthy subjects and rheumatoid arthritis were coupled with protein G sepharose beads and then further incubated with preparations of mimivirus proteins. Mimivirus proteins retained on the IgG-protein G beads were identified by mass spectrometric peptide sequencing after trypsin digestion. The major capsid protein L425 was found in all samples followed by the putative GMC type oxidoreductase R135 and core protein L410, which were found in seven from ten samples (Table 7). Interestingly, the most frequent mimivirus proteins recognized by human sera were surface proteins according to the surface biotinylation study (Shah *et. al.*, 2013). However, mimivirus collagens did not appear among the proteins recognized by the human sera. This absence may be linked to the fact that mimivirus collagens contain abundant lysine residues, thereby yielding very short tryptic peptides that remained below the detection range of mass spectrometric peptide sequencing. Nevertheless, the recognition of multiple mimivirus proteins by human sera confirmed the exposure of humans to mimivirus.

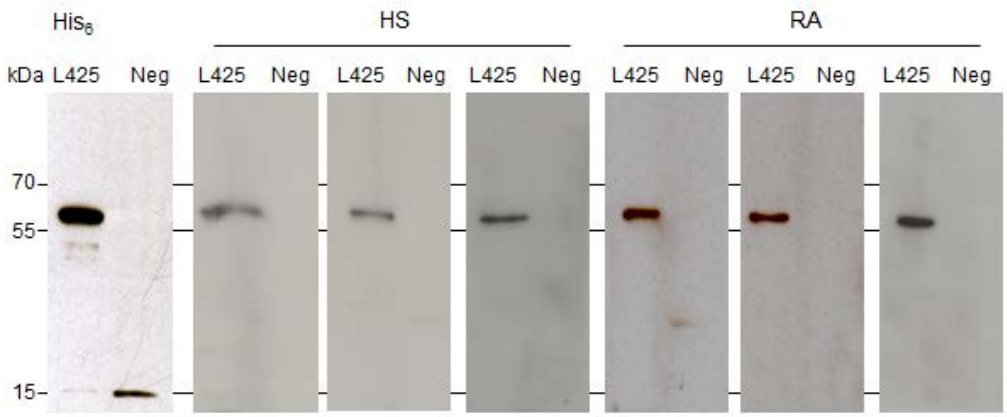
Table 7-Mimivirus proteins recognized by human serum IgG

ORF	**Protein annotation**	**Healthy subjects**					**Rheumatoid Arthritis**				
L425	Capsid protein	68	601	463	68	387	983	271	258	70	327
R135	Putative GMC-type oxidoreductase		370	394		153	774	255	163		161
L410	Core protein		223	284		179	685	123	133		97
R345	Putative regulator of chromosome condensation	52	123	145	28	58	350	185	77		217
R349	Uncharacterized protein	28	28	30	26	26	35	26		28	32

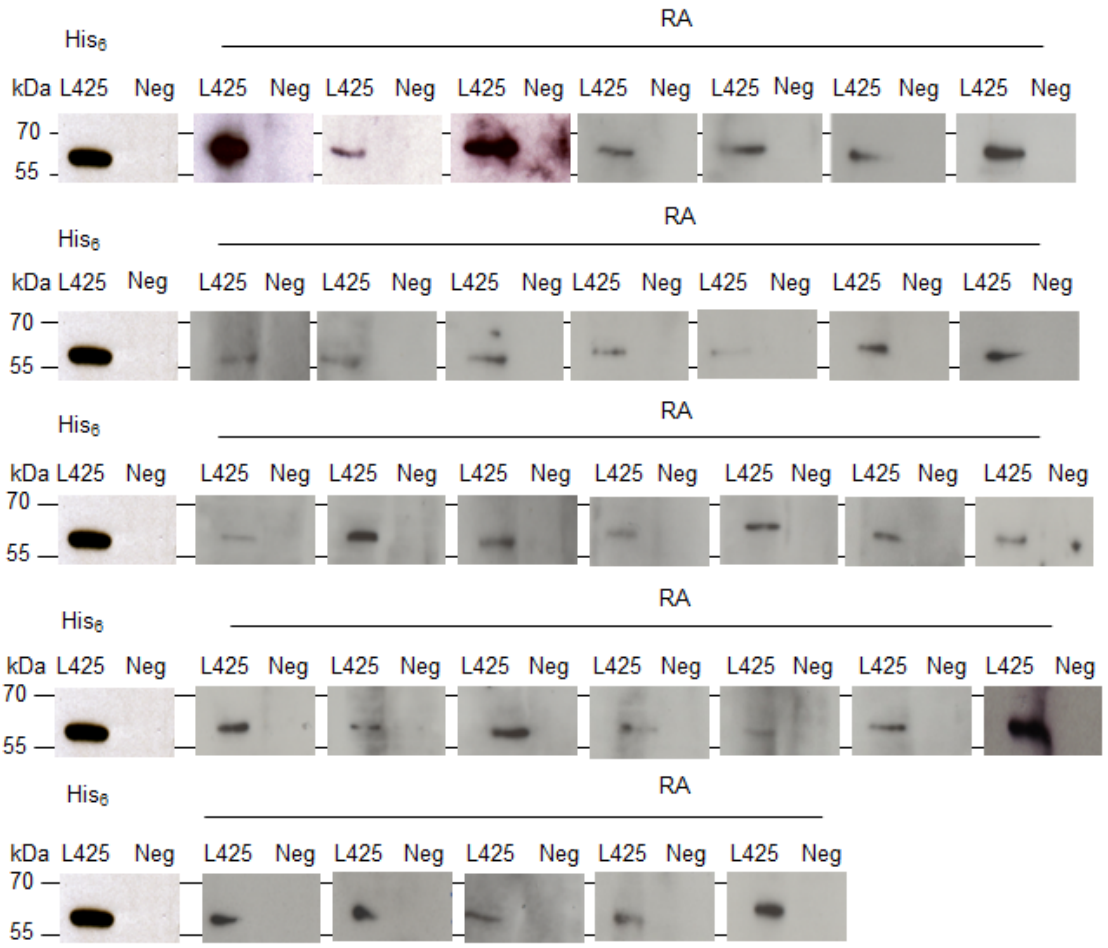
Scores for viral proteins recognized by 5 healthy subject sera and 5 rheumatoid arthritis sera are listed in columns for each serum tested. Values indicate Mascot scores representing the probability of positive matches for the recognized proteins. Scores above 25 were significant for $p<0.05$.

To further validate the occurrence of antibodies against specific mimivirus proteins in human sera were analysed by Western blot. The major capsid L425 and collagen L71 proteins were expressed as His_6-tagged recombinant proteins in *E. coli* and purified on Ni^{2+}-sepharose columns. Pools of 100 healthy subject sera and 100 rheumatoid arthritis sera were probed against the recombinant L425 and L71 mimivirus proteins. The reactivity of sera towards L425 was examined for three reasons, since L425 was detected among the mimivirus proteins captured by immunoprecipitation in the previous experiment, L425 is the major capsid protein thus is the most abundant protein in the mimivirus protein extract and L425 could be a prominent protein that the immune system recognises after viral entry into the host. For the 100 healthy subject and 100 rheumatoid arthritis sera tested, 30 and 36 sera respectively recognized the major capsid protein L425 (Fig. 24A and B). This results confirmed that exposure to mimivirus is common in the human population. The detection of IgG titers against the mimivirus capsid protein L425 in 30% of tested sera suggests repeated antigenic challenge is probably caused by repeated contact with mimivirus.

A


His6
HS
RA
kDa L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg
70
55
15

B


His6
RA
kDa L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg L425 Neg
70
55

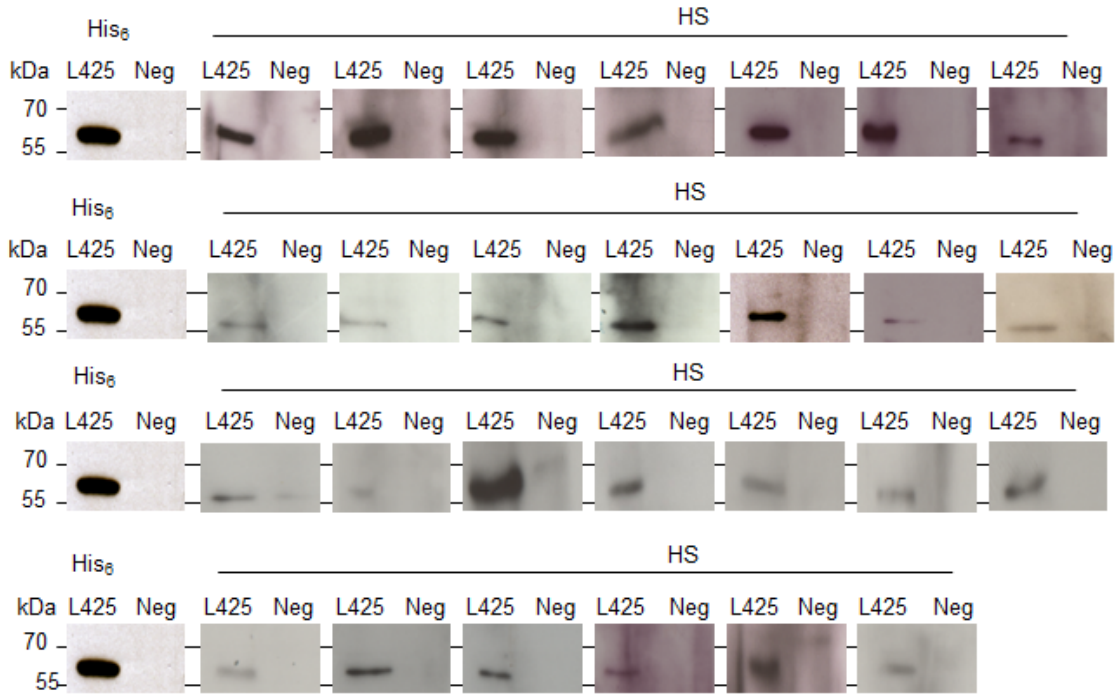
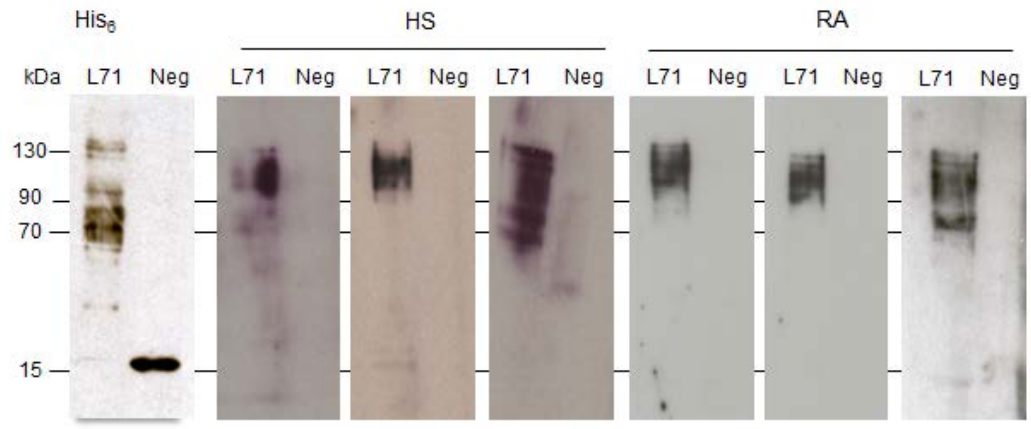


Fig. 24-Recognition of mimivirus capsid protein L425 by human sera. Pools of 100 healthy subjects (HS) and 100 rheumatoid arthritis (RA) patient's sera were tested for reactivity with mimivirus capsid protein L425 by Western blotting. 30 HS and 36 RA sera recognized L425. (**A**) Representative Western blots of 3 sera each from HS and RA patients recognizing mimivirus capsid protein L425. (**B**) The remaining Western blots of sera from 27 HS and 33 RA sera recognizing L425 are shown. The 70 HS and 64 RA sera which did not show reactivity to L425 are not shown. Sera were diluted 1:4000. A 15 kDa fragment of His_6-tagged human GLT25D2 protein was used as negative control (Neg). Positions of recombinant L425 and Neg proteins in the blots are shown at the left of each panel using an anti-His_6 antibody (His_6).

Even though L71 collagen protein was not captured by immunoprecipitation in the previous experiment, the reactivity of sera towards L71 collagen protein was examined for two reasons, since this protein is expressed at the surface of mimivirus particles (Shah *et. al.*, 2013) and is one of the seven collagen-like proteins encoded by the viral genome. Reactivity of human sera towards mimivirus collagen L71 was more discriminatory. Whereas only 6 healthy subject sera recognized the mimivirus collagen L71, 22 rheumatoid arthritis sera were positive for the mimivirus collagen L71 (Fig. 25A and B). Accordingly, this confirmed that the reactivity to mimivirus collagen was 3.5 times more frequent in the pool of rheumatoid arthritis sera in comparison to the limited reactivity of sera from healthy subjects.

A



B

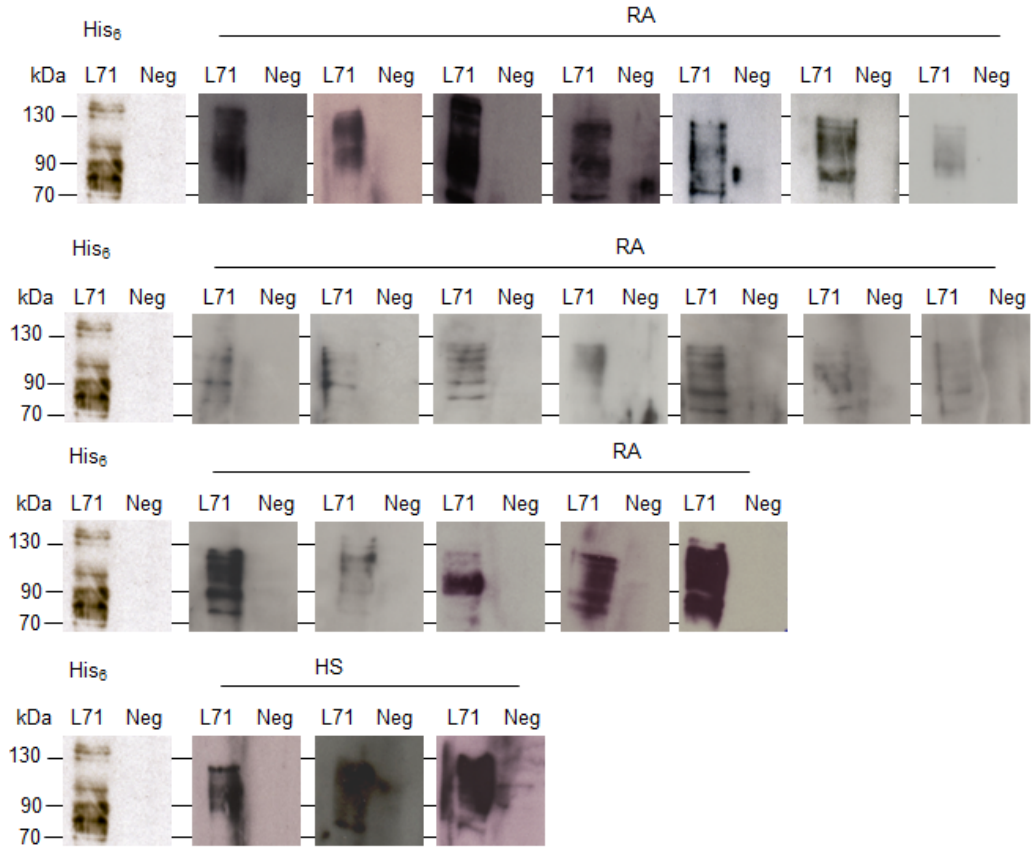
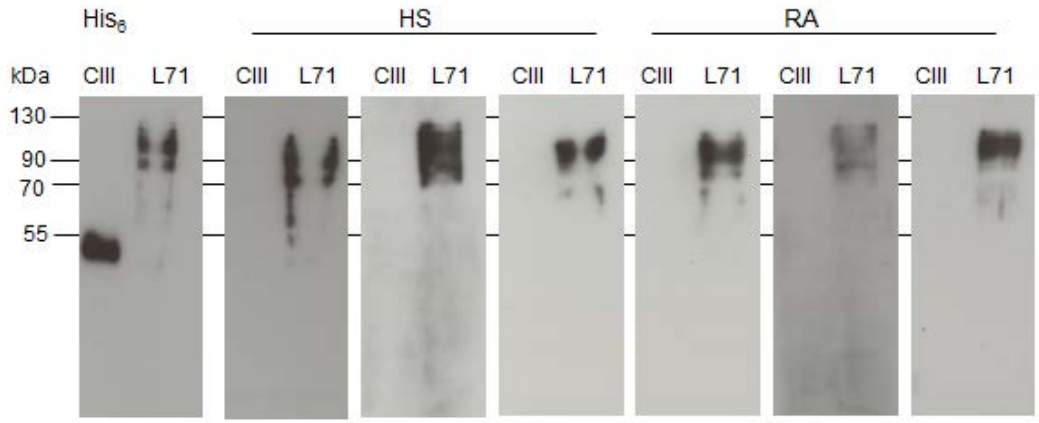


Fig. 25-Recognition of mimivirus collagen-like protein L71 by human sera. Pools of 100 healthy subjects (HS) and 100 rheumatoid arthritis (RA) patient's sera were tested for reactivity with mimivirus collagen-like protein L71 by Western blotting. 6 HS and 22 RA sera recognized L71. (**A**) Representative Western blots of 3 sera each from HS and RA patients recognizing mimivirus collagen-like protein L71. (**B**) The remaining Western blots of sera from 3 HS and 19 RA sera recognizing L71 are shown. The 94 HS and 78 RA sera which did not show reactivity to

L71 are not shown. Sera were diluted 1:4000. A 15 kDa fragment of His_6-tagged human GLT25D2 protein was used as negative control (Neg). Positions of recombinant L71 and Neg proteins in the blots are shown at the left of each panel using an anti-His_6 antibody (His_6).

Furthermore, one might argue that the human sera recognizing mimivirus collagen proteins might be simply due to cross-reactivity with several collagen-like proteins present in many bacteria existing maybe in the gut of humans. Therefore, to exclude unspecific cross-reactivity of human sera towards polypeptides containing $[G\text{-}X\text{-}Y]_n$ collagenous domains, the recognition of L71-positive sera were also tested for a fragment of human collagen type III encompassing 114 [G-X-Y] repeats and lacking N- and C-propeptides. None of the 28 human sera (6 healthy subjects + 22 rheumatoid arthritis) which were positive for mimivirus L71 recognize the 36 kDa $[G\text{-}X\text{-}Y]_{114}$ collagen type III construct (Fig. 26A and B). This clearly thereby demonstrates the specificity of the antibody response to mimivirus L71 collagen.

A



B

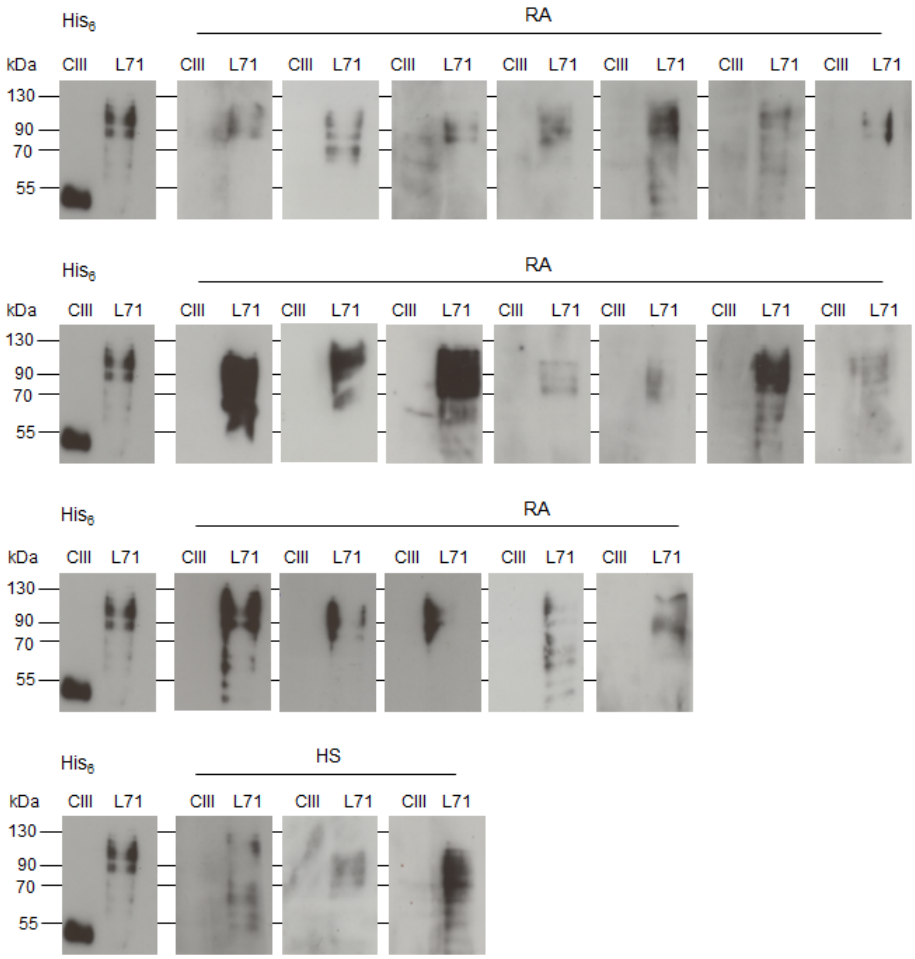


Fig. 26-Specific recognition of mimivirus collagen L71 by human sera. Western blots of L71-positive sera from healthy subjects (HS) and rheumatoid arthritis (RA) patients recognizing mimivirus collagen L71 but not a fragment of human collagen type III containing 114 [G-X-Y] repeats (CIII). (**A**) Representative Western blots of L71-positive sera from 3 sera each of HS and RA patient's recognizing mimivirus collagen L71 but not CIII. (**B**) The remaining blots of sera

from 3 HS and 19 RA sera recognizing mimivirus collagen L71 but not CIII are shown. Sera were diluted 1:4000. Positions of recombinant L71 and CIII proteins in the blots are shown at the left of the panel using an anti-His_6 antibody (His_6).

Additionally, one of the characteristic features of rheumatoid arthritis is the presence of elevated IgG titers against endogenous CII. Anti-CII titers were measured by ELISA in a subset of human sera which did not show reactivity to mimivirus collagen L71 in the Western blot analyses and in human sera which showed reactivity to mimivirus collagen L71 (Fig. 27). As expected, the rheumatoid arthritis sera (black symbols) in both groups showed elevated titers of IgG against endogenous CII. However the majority of the healthy subjects sera (red circles) tested in the L71 negative group showed no or less IgG titers against endogenous CII. Interestingly, 4 out of 6 healthy subjects sera (red squares) in the L71 positive group also showed high IgG titers against endogenous CII and this also correlates with them recognizing the mimivirus collagen L71 protein as well. This shows that majority of the human sera (healthy subjects and rheumatoid arthritis) recognizing mimivirus collagen L71 also has high titers against endogenous CII as well.

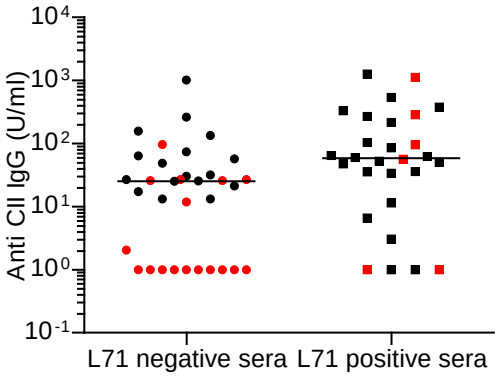


Fig. 27-Anti-collagen type II IgG titers in human sera. Levels of serum IgG measured by ELISA against endogenous human collagen type II (CII) in a subset of human sera (n=34) which did not recognize L71 in the Western blot analyses (L71 negative sera) and human sera (n=28) which recognized L71 in the Western blot analyses (L71 positive sera). Red symbols represent healthy subject's sera and black symbols represent rheumatoid arthritis patient's sera. Horizontal bars show medians.

Isolation of giant viruses from environment

Unicellular protist amoebae such as *Acanthamoeba* species are ubiquitously found in the natural environment especially in water and soil [214]. These amoebae feed on bacteria but however many bacterial species have established various strategies to overcome the phagocytosis and bactericidal activity of amoebae [215,216]. Some bacteria have developed mechanisms to produce toxins and pigments to avoid phagocytosis by amoebae [215,216]. Other bacteria have evolved in a way that they have become obligatory intracellular parasites of amoebae and are not able to grow outside amoebae [215,217]. The identification of the human pathogenic bacteria *Legionella pneumophila* associated with various species of amoebae suggested clinical microbiologists the presence of pathogens growing in amoebae to be present in natural environments [218-220]. Furthermore, now amoebae are termed as 'melting pot of evolution' due to identification of other pathogenic micro-organisms such as *Rickettsiaceae*, *Parachlamydiaceae* species which have adapted to become resistant to digestion by amoebae [47,48]. Additionally, the unexpected discovery of mimivirus from a cooling water tower which was initially thought to be a Gram-positive bacterium [1,3] led to growing interests to identify and characterize other amoeba-associated viruses such as mamavirus [25,45], megavirus [26], marseillevirus [30] and many more. The isolation of giant viruses from various natural environments in the recent past [54,55], led to the question whether giant viruses are also detectable in the local natural environments around Zurich. Therefore water samples from 14 various natural aquatic environments (lake, river, ponds and fountains) were collected and processed through a filtration protocol. The culture supernatants were then infected on flasks of amoeba *A. polyphaga* to observe if any of these water samples harboured any giant virus which could infect and propagate in amoeba. Interestingly, 3 out of 14 samples collected did show infectivity in amoeba i.e. amoeba were lysed in culture flasks (Table 8).

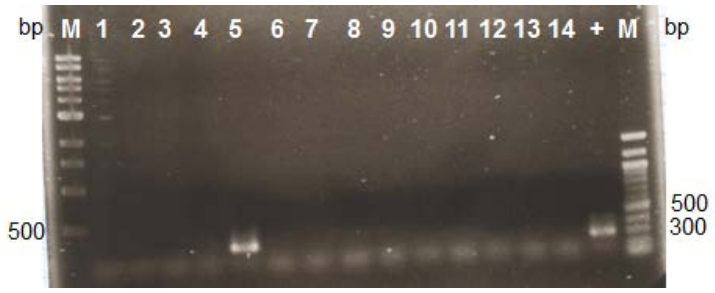
Table 8-Infectivity of amoeba by the various water samples (++=amoeba were infected and were lysed, --=amoeba were not infected)

Sample #	Collected from	Infectivity	Sample #	Collected from	Infectivity
1	Pond in forest, Opfikon	--	8	Greifensee	--
2	Biotope water, Opfikon	--	9	Furtbach, Otelfingen	++
3	Tessin lake	--	10	Artificial lake, Opfikon	--
4	Katzensee	--	11	Fountain water, Otelfingen	--
5	Zurich lake, Bellevue	++	12	Limmat	--
6	Fountain water, Rentenwiese	--	13	Irchel pond	--
7	Pond E floor, Irchel	--	14	Zurich lake, Rentenwiese	+

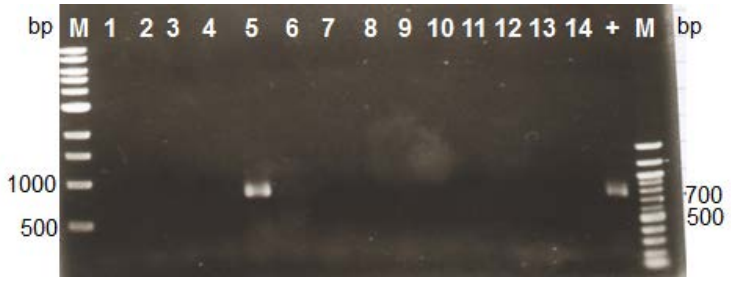
Further, from the culture supernatants of all 14 samples, DNA was extracted and PCR was performed with degenerative primers for the well conserved polymerase gene in mimivirus (PolB1 and PolB2) and marseillevirus (AVSpol) to see if any mimivirus-like, marseillevirus-like giant viruses are present in these water samples. Surprisingly, we observed that only sample # 5 (Zurich lake, Bellevue) showed a positive band with PolB1 and PolB2 giving the first hint that mimivirus-like giant virus may be present in the lake (Fig. 28A and B). The positive control for the PCR was DNA extracted from mimivirus particles which also showed the same positive band. But however, the other two water samples (# 9 and 14) showed infectivity in amoeba but no positive PCR product, so we postulated that some other giant virus infecting amoeba may be present in these two water samples. Thus, a marseillevirus polymerase (AVSpol) PCR was performed on all samples. Indeed, we observed that sample # 9 (Furtbach, Otelfingen) and sample # 14 (Zurich lake, Rentenwiese) showed a positive band with AVSpol

suggesting that marseillevirus-like giant virus may be present in these two water samples (Fig. 28C). Here of course the positive control did not show a band since DNA extracted from mimivirus was used as a positive control and this PCR was performed with marseillevirus specific primers. Sample # 14 showed positive PCR banding much later in the analysis and thus all further work was focussed more on sample # 5 and 9.

A



B



C

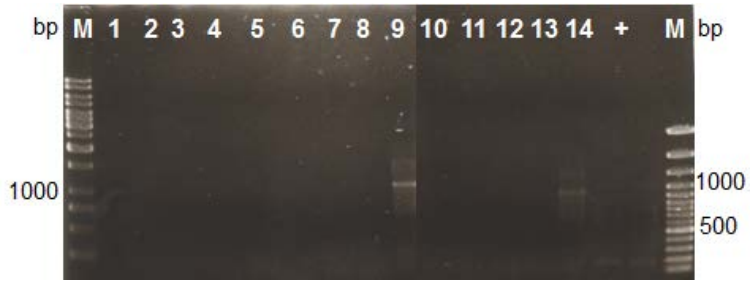


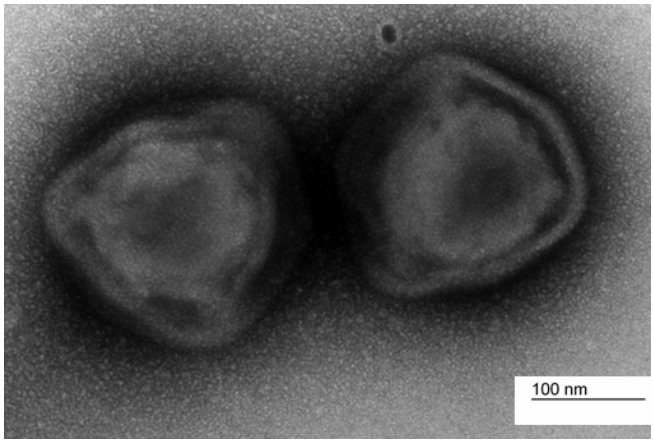
Fig. 28-PCR analyses for detection of giant viruses in environmental water samples. (Figure credits-S. Schneeberger) PCR was performed using various primers from DNA extracted from infected amoeba culture supernatants. (**A**) PCR with PolB1 primers and sample # 5 showing a positive band (**B**) PCR with PolB2 primers and sample # 5 showing a positive band (**C**) PCR with AVSpol primers and sample # 9 and 14 showing positive bands. M, DNA ladder; +, positive control (mimivirus); bp, base pairs.

Since PCR results gave us a first hint that giant viruses may be present in the water samples, the next step was to observe if these giant viruses we isolated have any morphological similarities with any known giant viruses. For this reason, the giant viruses from the two water samples (# 5 and 9) were harvested and were analysed by electron microscopy to observe the structure of the isolated organism infecting the

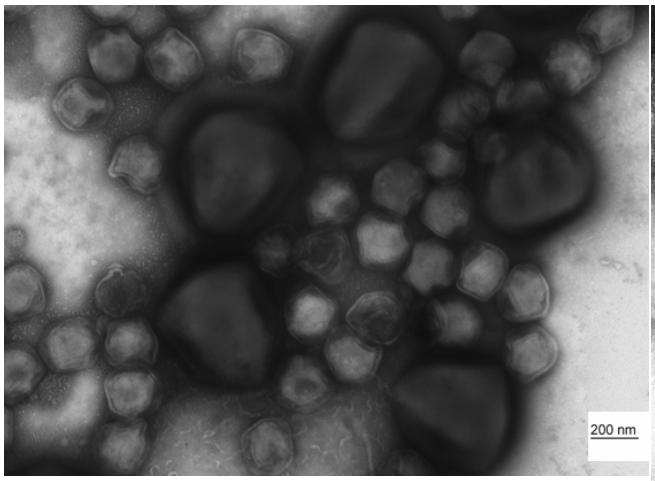
amoebae. From the electron micrographs we observed that from sample # 9 (Furtbach, Otelfingen), a virus with an icosahedral symmetry with central dense core with short fibres surrounding its capsid was isolated (Fig. 29A). This morphology highly resembles the structure of marseillevirus belonging to the putative new family of *Marseilleviridae* of the NCLDV group of viruses suggesting water sample # 9 harbours a marseillevirus-like isolate. This also correlates with the PCR results wherein sample # 9 showed positive band for marseillevirus specific primers.

Surprisingly, from sample # 5 (Zurich lake, Bellevue) there seems to be a mix of two giant viruses (Fig. 29B and C). One relatively smaller virus appears to be more like marseillevirus with its ~200 nm capsid diameter surrounded by small fibrils. The other relatively bigger giant virus (Fig. 29C) has a capsid diameter of ~ 500-600 nm with a central dense core, an icosahedral symmetry and long fibres attached to its surface. This morphology closely resembles the structure of mimivirus belonging to the *Mimiviridae* family of NCLDV group of viruses indicating presence of mix of two giant viruses in the Zurich lake. Correlating with the PCR results, sample # 5 also showed positive band for mimivirus specific primers suggesting the presence of mimivirus-like giant virus. This is an on-going experiment and currently work is on its way in the lab to separate these two viruses by serial dilution and immunoprecipitation protocols from sample # 5 so as to clearly identify the two viruses.

A



B



C

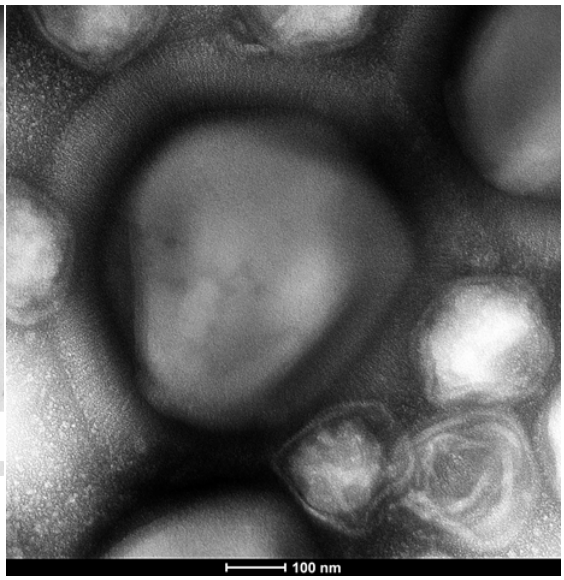
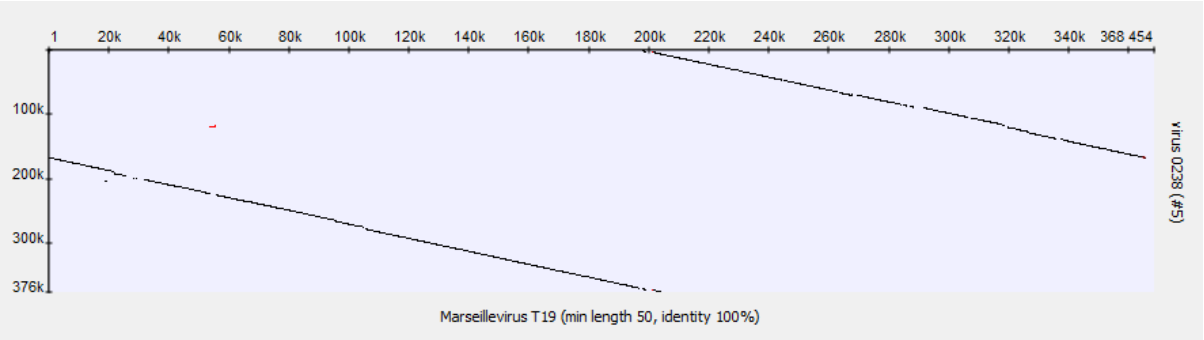


Fig. 29-Electron microscopy of virus isolates from water samples. (**A**) Electron micrograph of virus isolated from sample # 9 (Furtbach, Otelfingen). The capsid diameter is ~200 nm with a central dense core, icosahedral symmetry and short fibres around the capsid suggesting it to be marseillevirus-like isolate. (**B**) Electron micrograph of virus isolate from sample # 5 (Zurich lake, Bellevue). There seems to be a mix of two giant viruses, one relatively smaller ~200 nm in capsid diameter and other bigger virus ~ 500-600 nm in diameter. (**C**) Zoomed image of (**B**). The relatively bigger virus with its large capsid also showing icosahedral symmetry and long fibres attached to its capsid suggesting it to be mimivirus-like isolate.

Furthermore, the whole genomic DNA extracted from sample # 5 and 9 were also analysed by whole genome sequencing to see the best possible sequence matches. Two different contigs of genome sequences were obtained from sample # 5 and this also correlates with the electron microscopy data wherein two different viruses were observed. The smaller contig of 376,310 bp when compared with marseillevirus (T19 strain) sequence, we observed direct repeat matches between the two sequences but not a complete alignment and this could be because of recombination between the two viruses present in the sample (Fig. 30A).

A



B

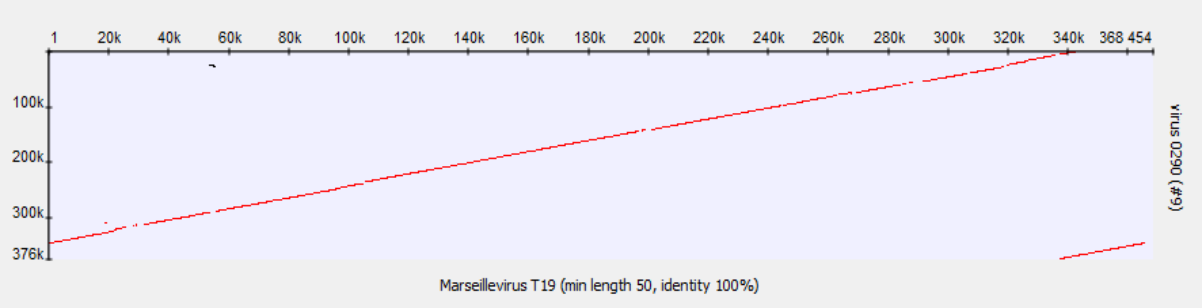


Fig. 30-Dotplot of sequence alignment of virus isolates with marseillevirus. Whole genome sequences retrieved were sequence aligned with genome sequence of marseillevirus (GenBank Accession number NC_013756) using the Unipro UGENE software. Sequences were matched with 100% similarity with a minimum repeat length of 50 bp. The marseillevirus sequence was plotted on the horizontal X axis and the virus isolate sequences on the vertical Y axis of the plot. Red indicates inverted repeats and black indicates direct repeats. The diagonal line represents continuous matches (or repeats) between the two sequences. (**A**) Dotplot with virus isolate from sample # 5 with marseillevirus showing sequence alignment with two diagonal black lines suggesting matched sequences. (**B**) Dotplot with virus isolate from sample # 9 with marseillevirus showing a red diagonal line suggesting over 95% matched sequences but however in an inverted repeat.

At this stage however, to comment on the relatively larger virus from sample # 5 is rather difficult since only when the two viruses are distinctly separated we can clearly identify the virus isolates. However for sample # 9, the genome sequence obtained contained 376,151 bp which showed more than 95% sequence similarity with 368,454 bp of marseillevirus (T19 strain) and has 7697 bp more than marseillevirus. The 95% matched sequences suggested the virus isolate closely resembles marseillevirus but however this match was in an inverted repeat fashion (Fig. 30B). When the two dotplots are compared to each other it seems evident that the virus isolate from sample # 9 and the relatively smaller virus isolate from sample # 5 are clearly different viruses but most probably belonging to the putative *Marseilleviridae* family of NCLDV group.

Thus in totality the task to possibly isolate and characterize giant viruses from natural environments is not entirely complete but however the work so far interestingly gives us a hint of the presence of possibly *Mimiviridae* and *Marseilleviridae* members in the water resources around Zurich.

Discussion

The present study demonstrated that mice immunized with mimivirus proteins including viral collagens generated auto-reactive anti-collagen antibodies. Additionally, an auto-reactive T cell response and physical abnormalities of bone and cartilage destruction in joint articular tissues was observed in immunized mice. A possible relationship between exposure to mimivirus collagen and the development of auto-immunity was supported by the presence of IgG against mimivirus collagen among rheumatoid arthritis patients. These findings suggests and gives a first hint that repeated exposure to mimivirus can lead to the formation of antibodies against virus collagen and to a breakage of immune tolerance of endogenous collagens.

Firstly, using the CIA mouse model which closely resembles rheumatoid arthritis in humans [187,188], we showed that after intra-dermal immunization with mimivirus proteins mice did develop a phenotype consistent with limb inflammation and swelling. Although a fascinating observation, this phenotype was not as severe as that observed in mice immunized with bovine CII. Histologically, the cartilage lining and bone structure of mimivirus proteins immunized mice were in the initial phases of destruction and clear signs of synovial hyperplasia were observed. Infiltration of cells in the synovial cavity was also observed. Immunohistochemistry could be used in the future to identify and pin-point the infiltrating cell type, which is likely to be either macrophages or activated T cells or infiltrating leukocytes. Inflamed synovium results from the action of proinflammatory cytokines such as TNF, IL 1, IL 6 and matrix degrading enzymes and molecules such as MMPs and RANKL which are secreted by the infiltrating cells [198-200]. Bead based immunoassays followed by flow cytometry of sera samples from mice could be used to confirm the up-regulation of proinflammatory cytokines in immunized mice. Nevertheless, using an *in vivo* fluorescence imaging technique the presence of inflammation associated proteases such as cathepsins were observed in the limbs of immunized mice. This observation confirmed the presence of cathepsins in joint articular tissues of mice as is generally observed during arthritis [199,200], a process which in turn can assist in the recruitment of osteoclast cells that contribute to bone erosion and destroy collagen matrices. Furthermore, breakage of immune tolerance to collagen was observed when high IgG titers against self-

endogenous CII were detected in immunized mice. High titers of IgG against CII are frequently observed and considered a hallmark of CIA and rheumatoid arthritis [196,197]. This result is quite interesting as in the past studies demonstrated that mice deficient in B cells do not produce autoantibodies to CII and thus no do not develop CIA [221]. This supports the role of autoantibodies against endogenous CII in the development of arthritis, even after immunization with mimivirus proteins. In addition to a humoral immune response, T cell auto-reactivity was also observed in recall assays when cells isolated from draining lymph nodes of immunized mice were seen to proliferate *in vitro* when stimulated with mouse, bovine and mimivirus L71 collagenous fragments. This observation is quite interesting considering that a fragment of mimivirus collagen L71 shares 73% sequence similarity with one of the major immunodominant T cell epitopes of CII, which is highly recognized in rheumatoid arthritis [85,86]. This shows that APCs are capable of presenting self and mimivirus collagenous peptides with the appropriate MHC molecule to T cells under the co-stimulatory signals of B cells and thereby initiating an autoimmune response. In a previous study, the pathogenic potential of mimivirus in a mammalian host was examined by inoculating infectious viral particles via an intra-cardiac route in mice [59]. The inoculated mice developed acute atypical pneumonia and it was shown that mimivirus could be re-cultured from lung homogenates of inoculated mice [59]. However, observations in the present study using an established mouse model of CIA wherein mice were immunized with viral proteins, clearly show that on repeated exposure to mimivirus potentially by molecular mimicry, mimivirus is capable to trigger an autoimmune response possibly to collagens.

Further, taking advantage of human blood sera samples a human perspective on rheumatoid arthritis was also studied here. The main factors involved in the pathogenicity of rheumatoid arthritis could either be genetic or environmental factors. Rheumatoid arthritis is an autoimmune disease with a significant environmental component as supported by twin studies [145,222,223]. Repeated contact with collagen antigens found in the environment may promote the development of cross-reactive anti-collagen antibodies and lead to inflammation in collagen rich tissues. As shown in this study, anti-mimivirus IgG and IgM responses were detected not only in disease conditions but also in healthy subjects. Based on this finding it can be said that infection

might be common but not autoimmunity. Since mimivirus is routinely found in the natural environments humans are at a constant risk of exposure to mimivirus and could trigger an immune response against the virus. However, it may only be in some genetically predisposed individuals that this environmental exposure triggers an autoimmune response, leading to the clinical appearance of rheumatoid arthritis. This can also be correlated to the fact that exposure to mimivirus was confirmed when 30% of healthy subjects and 36% of rheumatoid arthritis sera recognized the mimivirus major capsid protein L425. This suggests that the host immune system recognizes predominantly the major capsid protein L425 upon viral entry and based on this recognition thus mounts an immune response. However, the 3.5 times more frequent recognition of mimivirus collagen L71 by rheumatoid arthritis sera than healthy subjects shows the potential antigenicity of viral collagen proteins. Antibodies against collagens can either recognize the triple helical conformation or peptides sequences in the triple helical domain or in telopeptides [224]. The specificity of antibodies against collagen depends on the activation of a humoral response, or on a combination of cell-mediated and humoral responses [224]. The reactivity of human sera to mimivirus collagen L71 indicates that epitopes are recognized based on their amino acid sequence and not their 3D conformations. This notion was supported by finding no reactivity of the L71-positive human sera for the human collagen type III $[G\text{-}X\text{-}Y]_{114}$ polypeptide used as negative control. The lack of recognition of the $[G\text{-}X\text{-}Y]_{114}$ polypeptide also indicated that the reactivity towards L71 was specific to mimivirus exposure and not the result of cross-reactivity to collagen-domain containing proteins such as those found in some Gram-positive bacteria.

Should we worry about mimivirus?

The serendipitous discovery and further sequencing of the mimivirus genome brought to light the unusual nature of genes and proteins present in an organism defined and classified as a 'virus' [3,4]. Mimivirus possess genes such as aminoacyl-tRNA synthetases, translation initiation and elongation factors and peptide chain release factors, which have never been identified in a virus and to date are known to be hallmarks of cellular organisms [3]. This occurrence of genes is not restricted to mimivirus but is also present in other giant viruses belonging to the same lineage of the

Mimiviridae family of NCLDVs, which have been recently isolated and characterized, such as mamavirus [25,45], megavirus [26], moumouvirus [44] and many more. The presence of unusual genes never seen in a virus before, led many scientists to believe that viruses, especially giant viruses, should be grouped as a fourth domain of life after archaea, bacteria and eukarya [27,52]. These giant viruses propagate in the ubiquitously present amoebae such as *Acanthamoeba* species. Interestingly, giant viruses are routinely found in natural aquatic, marine and soil environments suggesting that humans are at a constant risk of coming in contact with these viruses.

Mimivirus is most likely ingested through drinking water and taken up by dendritic cells and macrophages lining the gastrointestinal mucosa. Alternatively, virus particles may enter the airways as aerosols and be taken up by alveolar macrophages. In fact, mimivirus can be phagocytized by human and mouse macrophages although the virus cannot replicate in these cells [63]. This suggests that mimivirus peptides can be presented by APCs, which could trigger an appropriate immune response against the virus. Similarly, by seroconversion to mimivirus, infection has been linked to pneumonia in individual case studies [57,58] and also in groups of community and hospital acquired pneumonia [2,56,57] all without evidence for virus particles in disease cases. Moreover, conflicting results from groups have also shown no correlation between mimivirus and pneumonia patients, suggesting mimivirus is not a common respiratory pathogen [60,61]. However, this interspersed pathogenicity does not preclude a more general effect of mimivirus in priming an autoimmune response.

As explored in the present study, mimivirus possesses some proteins with comparable sequence similarity to mammalian proteins such as collagen-like proteins, thereby creating the possibility of antibody cross-reaction with endogenous proteins after viral exposure. Indeed, as shown here mimivirus has the potential to trigger an autoimmune response similar to arthritis in mice and human antibodies specifically cross-reacted with mimivirus collagen proteins. Thus humans coming in contact with mimivirus, could be at risk of compromised immune tolerance to endogenous collagens, which may trigger autoimmunity, via molecular/antigen mimicry. Considered together these studies suggest that, mimivirus can have the potential to be an emerging human pathogen. Thus the honest answer to the question would be ‘maybe’ but significant

work needs to be done in the field before the general public is alarmed by the presence of a giant virus in the environment.

Microbes and molecular mimicry

The body's immune system, which is often referred to as the 'sixth sense' of body, has the ability to distinguish between endogenous and foreign antigens and is required for both protection against autoimmune destruction of endogenous antigens and host defence against infectious foreign antigens [89]. Central and peripheral tolerance mechanisms such as deletion (negative selection), anergy, receptor editing, inhibition and suppression all exist to govern the tolerance mechanisms of B and T cells and thereby attempt to eliminate any auto-reactive lymphocytes [89-94]. The breakdown of any of these tolerance mechanisms results in the production of autoantibodies and an auto-reactive T cell response, leading to the clinical manifestation of many autoimmune diseases. Infections with foreign agents such as bacteria and viruses have long been postulated in generation of autoimmunity. Although the precise mechanisms by which these infectious agents trigger the generation of autoantibodies remains unclear, many studies have revealed cross-reactivity between endogenous proteins and foreign-antigens (i.e. molecular mimicry playing an important role) [225].

The most common and classical examples of autoimmune diseases linked to molecular mimicry due to acute infections of distinct bacteria are Guillain-Barré syndrome and rheumatic fever. The major characteristic feature of Guillain-Barré syndrome is demyelination in the peripheral nervous system, typically with the onset of sudden limb weakness with lymphocytic infiltration [119,120]. Infections with *C. jejuni*, which is the leading cause of gastroenteritis, have been associated with the development of the syndrome and are well documented not only by serological analysis but also by isolation of the bacterium from patients [226]. In a previous study, it was shown that only 2% of healthy controls and 26% of Guillain-Barré syndrome patients showed the presence of *C. jejuni* [226]. Here, molecular mimicry occurs when antibodies against *C. jejuni* lipooligosaccharides cross-react with endogenous GM1 gangliosides on nerve cells as the outer polysaccharide moieties of lipooligosaccharides of *C. jejuni* show striking structural similarities with gangliosides on nerve cells [119,120]. In the clinical course of the disease, the presence of high titers of antibodies to GM1 gangliosides are

detected and motor neuron functions are impaired [226]. Another case of molecular mimicry with group A streptococcus infections is associated with rheumatic fever [227,228]. The M protein of streptococcus has been shown to play a major role in the development of cross-reacting antibodies [228]. The M protein which has an extended alpha helical structure bears major sequence similarity with many human proteins such as myosin, tropomyosin, laminin and keratin [228]. The most serious consequence of rheumatic fever is severe damage to heart valves [228]. Indeed, studies have shown that murine and human antibodies specific to streptococcus M protein also cross-react with cardiac myosin due to a stretch of amino acid sequence similarity between the two proteins [227,228]. However, since different strains of streptococcus colonize mucus membranes and skin, rheumatic fever is generally caused by streptococcal pharyngitis but not skin streptococcal infections [227,228].

Association of autoimmunity with molecular mimicry is not restricted to pathogenic bacterial infections but has also been shown with commonly occurring viral infections. One common example is Epstein-Barr virus, which causes several lymphoproliferative diseases such as acute infectious mononucleosis, but has also been linked with autoimmune systemic lupus erythematosus (or simply lupus) [229]. Molecular mimicry was shown to play an important role when the humoral response against Epstein-Barr virus, particularly against Epstein-Barr virus nuclear antigen 1, was different between lupus patients and healthy controls [229,230]. Antibodies raised against Epstein-Barr virus nuclear antigen 1 were shown to cross-react with lupus specific auto-antigens such as Ro and Sm (RNA binding nuclear proteins) due to structural similarities [230-232]. Furthermore, a transgenic mouse model has been used to show that B cells expressing Epstein-Barr virus encoding latent membrane protein 2A evade the self-tolerance mechanisms, thereby developing autoimmunity [233]. The latent protein stimulates hypersensitivity to Toll-like receptors, further leading to anti-Sm B cells and triggering the production of anti-Sm autoantibodies [233]. Further, Epstein-Barr virus infected B cells were seen to proliferate and be resistant to apoptosis as they show the presence of virally encoded anti-apoptotic molecules [234]. This leads to co-stimulatory signals that activate auto-reactive T cells, which help to produce cytokines, ultimately resulting in target organ damage (i.e. skin lesions and inflammation) [234]. Interestingly, Epstein-Barr virus infection has also been linked to rheumatoid arthritis

in some cases but causality however has not yet been proven [235]. Epstein-Barr virus infection was proposed to contribute to citrullination of proteins, thereby producing anti-cyclic citrullinated peptide antibodies and thus was associated with rheumatoid arthritis [235]. Another viral infection that causes severe liver damage, Hepatitis B virus, has been linked to autoimmune hepatitis. Viral infection leads to increased antibody production by B cells and effector cytotoxic T cells, which help in eliminating the circulating viral particles in the host, thus preventing (re)infection [236]. However, these T cells can also cause severe liver damage via perforin-mediated cytotoxicity and Fas-mediated apoptosis [236,237]. In fact, four of the human nuclear proteins involved in structural functions and two smooth muscle proteins involved in muscle contraction show sequence similarity with Hepatitis B virus DNA polymerase [238]. Autoimmunity was demonstrated to be a common feature with Hepatitis B viral infection, as many patients with chronic viral infection have been found to be seropositive for autoantibodies against nuclear and smooth muscle proteins [238]. Cross-reactivity between viral and host proteins during a normal immune response to infection could further deteriorate the clinical condition, exasperating autoimmune hepatitis [238,239]. Likewise, in this study exploring the phenomenon of molecular mimicry between mimivirus collagens and mammalian collagens, we postulate a possible environmental link associated with rheumatoid arthritis. This was shown by the potential of mimivirus to initiate an autoimmune arthritic response in mice and also by the frequent occurrence of antibodies recognizing mimivirus collagen in rheumatoid arthritis sera. Even though mimivirus has yet to be shown to replicate and proliferate in mammalian cells, repeated exposure to structurally similar proteins which are surface localized could lead to an autoimmune response.

Being multifactorial in nature, autoimmune diseases are like a jigsaw puzzle wherein several genetic, environmental and stochastic (chance) factors form the different pieces of the puzzle which when comes together forms a complete picture. As seen above, environmental factors, in particular infection with bacteria or viruses by the process of molecular mimicry form a very important piece of the puzzle. In the future, molecular mimicry may have important clinical implications for vaccination strategies. However, non-responsiveness to these vaccines can be associated to be the function of tolerance [118]. By the same argument, in genetically predisposed individuals, vaccination against

these infectious agents could activate pathways of molecular mimicry, resulting in contrary reactions to vaccines [118]. Research still needs to be conducted to avoid such consequences in a clinical setting.

However, in all these molecular mimicry studies it is very important to note that the bacterial or viral infection should not be considered as the sole causative agent of the corresponding autoimmune response. Molecular mimicry has been postulated in many autoimmune diseases solely on the basis of circumstantial evidence. Nevertheless, molecular mimicry could provide a physiological link between the existing knowledge on host defence and autoimmunity.

Future directions

In the present study, using the established CIA model for rheumatoid arthritis, mice were immunized with protein extracts from mature, whole mimivirus particles and an autoimmune response similar to arthritis was observed. To further support the hypothesis that mimivirus can trigger an arthritic like response in mice and indeed is rather collagen specific, mice can be immunized using only purified mimivirus collagen protein (like L71). To do so, large quantities of recombinant purified mimivirus L71 protein would be required since estimating how much protein would be needed to observe the matching phenotypic manifestation and immune response is rather difficult (currently on-going). Alternatively, if collagen in its native form can be extracted from mimivirus particles by pepsin digestions, this could also be useful for mice immunizations. As mentioned earlier, the analysis of up-regulation of specific infiltrating immune cells and proinflammatory cytokine responses in synovial joints in immunized mice would be an appealing experiment.

Similar to the immunodominant CII peptide in rheumatoid arthritis, the mimivirus collagen L71 peptide also contains two lysine residues. Additionally, mimivirus also harbours a novel bi-functional enzyme capable of hydroxylating the lysine residues and further glycosylating them in a unique fashion i.e. adding glucose directly rather than the conserved collagen glycosylation pattern of galactose-glucose [84]. Since previous studies have shown that T cells can be differentially activated with glycosylated or non-glycosylated form of the immunodominant CII peptide [85,86], it would be of interest to

investigate this phenomenon in this model as well to understand how T cells respond to either the glycosylated or non-glycosylated L71 peptide.

In this study, the negative control group of mice were immunized with PBS emulsified with Freund's adjuvant, which was mainly to have an adjuvant control so as to observe whether the immune response is antigen specific or adjuvant specific. No phenotypic observation or autoimmune response was observed in this group of mice, suggesting that the immune response is clearly antigen specific. However, another important negative control group for future studies would be to use mice immunized with an unrelated virus such as any member of NCLDV group or a giant virus, lacking collagen-like proteins such as marseillevirus, to further confirm the specificity of the reaction induced by mimivirus (currently on-going).

Since similar immunizations in humans cannot be done with mimivirus proteins as was performed here with mice, we have to rely on experiments performed on biological samples obtained from rheumatoid arthritis patients. If it were to be possible to obtain B and T cells from rheumatoid arthritis patients, a very captivating experiment would be to test for the proliferation of human T cells upon stimulation with mimivirus proteins under the co-stimulatory signals from B cells and other released cytokines. This experiment would serve to further confirm the hypothesis in humans too. Furthermore, since rheumatoid arthritis is a multifactorial autoimmune disease, it would also be interesting to perform genetic analyses and observe if some genetically predisposed individuals are at a higher risk of acquiring the disease in the presence of an environmental trigger such as mimivirus infection.

Conclusions

In conclusion, in light of the structural similarity between mimivirus collagen and human collagen sequences, the present study proposes that giant viruses expressing collagen represent a potential environmental risk factor contributing to the development of rheumatoid arthritis. A systematic survey of mimivirus distribution in the environment will contribute to a better appreciation of the environmental risk associated to such giant viruses in relation to the geographical incidence of rheumatoid arthritis.

References

1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, et al. (2003) A giant virus in amoebae. Science 299: 2033.
2. Raoult D, La Scola B, Birtles R (2007) The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent. Clin Infect Dis 45: 95-102.
3. Raoult D, Audic S, Robert C, Abergel C, Renesto P, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306: 1344-1350.
4. Iyer LA, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 117: 156-184.
5. Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 75: 11720-11734.
6. Raoult D, Forterre P (2008) Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol 6: 315-319.
7. Koonin EV, Yutin N (2010) Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology 53: 284-292.
8. Yutin N, Koonin EV (2009) Evolution of DNA ligases of nucleo-cytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity. Biol Direct 4: 51.
9. Schramm B, Locker JK (2005) Cytoplasmic organization of POXvirus DNA replication. Traffic 6: 839-846.
10. Katsafanas GC, Moss B (2007) Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2: 221-228.
11. Garcia-Beato R, Salas ML, Vinuela E, Salas J (1992) Role of the host cell nucleus in the replication of African swine fever virus DNA. Virology 188: 637-649.
12. Ogata H, Toyoda K, Tomaru Y, Nakayama N, Shirai Y, et al. (2009) Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. Virol J 6: 178.
13. Federici BA, Bideshi DK, Tan Y, Spears T, Bigot Y (2009) Ascoviruses: superb manipulators of apoptosis for viral replication and transmission. Curr Top Microbiol Immunol 328: 171-196.
14. Bideshi DK, Demattei MV, Rouleux-Bonnin F, Stasiak K, Tan Y, et al. (2006) Genomic sequence of Spodoptera frugiperda Ascovirus 1a, an enveloped, double-stranded DNA insect virus that manipulates apoptosis for viral reproduction. J Virol 80: 11791-11805.
15. Chinchar VG, Hyatt A, Miyazaki T, Williams T (2009) Family Iridoviridae: poor viral relations no longer. Curr Top Microbiol Immunol 328: 123-170.
16. Wilson WH, Van Etten JL, Allen MJ (2009) The Phycodnaviridae: the story of how tiny giants rule the world. Curr Top Microbiol Immunol 328: 1-42.
17. Van Etten JL (2003) Unusual life style of giant chlorella viruses. Annu Rev Genet 37: 153-195.
18. Doms RW, Lamb RA, Rose JK, Helenius A (1993) Folding and assembly of viral membrane proteins. Virology 193: 545-562.
19. Olofsson S, Hansen JE (1998) Host cell glycosylation of viral glycoproteins--a battlefield for host defence and viral resistance. Scand J Infect Dis 30: 435-440.
20. Graves MV, Bernadt CT, Cerny R, Van Etten JL (2001) Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. Virology 285: 332-345.
21. Lippincott-Schwartz J, Yuan LC, Bonifacino JS, Klausner RD (1989) Rapid redistribution of Golgi proteins into the ER in cells treated with brefeldin A: evidence for membrane cycling from Golgi to ER. Cell 56: 801-813.
22. Que Q, Li Y, Wang IN, Lane LC, Chaney WG, et al. (1994) Protein glycosylation and myristylation in Chlorella virus PBCV-1 and its antigenic variants. Virology 203: 320-327.

23. Wang IN, Li Y, Que Q, Bhattacharya M, Lane LC, et al. (1993) Evidence for virus-encoded glycosylation specificity. Proc Natl Acad Sci U S A 90: 3840-3844.
24. DeAngelis PL (1999) Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. Cell Mol Life Sci 56: 670-682.
25. Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, et al. (2011) Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. Genome Biol Evol 3: 737-742.
26. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A 108: 17486-17491.
27. Nasir A, Kim KM, Caetano-Anolles G (2012) Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol 12: 156.
28. Van Etten JL, Lane LC, Dunigan DD (2010) DNA viruses: the really big ones (giruses). Annu Rev Microbiol 64: 83-99.
29. Monier A, Claverie JM, Ogata H (2008) Taxonomic distribution of large DNA viruses in the sea. Genome Biol 9: R106.
30. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, et al. (2009) Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106: 21848-21853.
31. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, et al. (2011) Lausannevirus, a giant amoebal virus encoding histone doublets. Environ Microbiol 13: 1454-1466.
32. Colson P, Pagnier I, Yoosuf N, Fournous G, La Scola B, et al. (2013) "Marseilleviridae", a new family of giant viruses infecting amoebae. Arch Virol 158: 915-920.
33. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, et al. (2003) Bacteriophage T4 genome. Microbiol Mol Biol Rev 67: 86-156, table of contents.
34. Virgin HWt, Latreille P, Wamsley P, Hallsworth K, Weck KE, et al. (1997) Complete sequence and genomic analysis of murine gammaherpesvirus 68. J Virol 71: 5894-5904.
35. Van Etten JL, Meints RH (1999) Giant viruses infecting algae. Annu Rev Microbiol 53: 447-494.
36. Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem 70: 369-413.
37. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, et al. (2009) Structural studies of the giant mimivirus. PLoS Biol 7: e92.
38. Klose T, Kuznetsov YG, Xiao C, Sun S, McPherson A, et al. (2010) The three-dimensional structure of Mimivirus. Intervirology 53: 268-273.
39. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, et al. (2008) Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. PLoS Biol 6: e114.
40. Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, et al. (2006) Mimivirus and the emerging concept of "giant" virus. Virus Res 117: 133-144.
41. Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D (2007) Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. PLoS One 2: e328.
42. Mutsafi Y, Zauberman N, Sabanay I, Minsky A (2010) Vaccinia-like cytoplasmic replication of the giant Mimivirus. Proc Natl Acad Sci U S A 107: 5978-5982.
43. Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, et al. (2006) Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J Virol 80: 11678-11685.
44. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, et al. (2012) Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol Evol 4: 1324-1330.

45. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, et al. (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455: 100-104.
46. Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A 107: 19508-19513.
47. Moliner C, Fournier PE, Raoult D (2010) Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev 34: 281-294.
48. Greub G, Raoult D (2004) Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17: 413-433.
49. Moreira D, Brochier-Armanet C (2008) Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol Biol 8: 12.
50. Filee J, Siguier P, Chandler M (2007) I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. Trends Genet 23: 10-15.
51. Filee J, Pouget N, Chandler M (2008) Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. BMC Evol Biol 8: 320.
52. Claverie JM, Ogata H (2009) Ten good reasons not to exclude giruses from the evolutionary picture. Nat Rev Microbiol 7: 615; author reply 615.
53. Moreira D, Lopez-Garcia P (2009) Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol 7: 306-311.
54. La Scola B, Campocasso A, N'Dong R, Fournous G, Barrassi L, et al. (2010) Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology 53: 344-353.
55. Boughalmi M, Saadi H, Pagnier I, Colson P, Fournous G, et al. (2012) High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. Environ Microbiol.
56. Berger P, Papazian L, Drancourt M, La Scola B, Auffray JP, et al. (2006) Ameba-associated microorganisms and diagnosis of nosocomial pneumonia. Emerg Infect Dis 12: 248-255.
57. La Scola B, Marrie TJ, Auffray JP, Raoult D (2005) Mimivirus in pneumonia patients. Emerg Infect Dis 11: 449-452.
58. Raoult D, Renesto P, Brouqui P (2006) Laboratory infection of a technician by mimivirus. Ann Intern Med 144: 702-703.
59. Khan M, La Scola B, Lepidi H, Raoult D (2007) Pneumonia in mice inoculated experimentally with Acanthamoeba polyphaga mimivirus. Microb Pathog 42: 56-61.
60. Dare RK, Chittaganpitch M, Erdman DD (2008) Screening pneumonia patients for mimivirus. Emerg Infect Dis 14: 465-467.
61. Larcher C, Jeller V, Fischer H, Huemer HP (2006) Prevalence of respiratory viruses, including newly identified viruses, in hospitalised children in Austria. Eur J Clin Microbiol Infect Dis 25: 681-686.
62. Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, et al. (2013) Evidence of the megavirome in humans. J Clin Virol.
63. Ghigo E, Kartenbeck J, Lien P, Pelkmans L, Capo C, et al. (2008) Ameobal pathogen mimivirus infects macrophages through phagocytosis. PLoS Pathog 4: e1000087.
64. Steenbergen JN, Shuman HA, Casadevall A (2001) Cryptococcus neoformans interactions with amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in macrophages. Proc Natl Acad Sci U S A 98: 15245-15250.
65. Gelse K, Poschl E, Aigner T (2003) Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev 55: 1531-1546.
66. Myllyharju J, Kivirikko KI (2001) Collagens and collagen-related diseases. Ann Med 33: 7-21.
67. Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20: 33-43.
68. Risteli M, Niemitalo O, Lankinen H, Juffer AH, Myllyla R (2004) Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. J Biol Chem 279: 37535-37543.

69. Mizuno K, Hayashi T, Bachinger HP (2003) Hydroxylation-induced stabilization of the collagen triple helix. Further characterization of peptides with 4(R)-hydroxyproline in the Xaa position. J Biol Chem 278: 32373-32379.
70. Berg RA, Prockop DJ (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem Biophys Res Commun 52: 115-120.
71. Spiro RG, Fukushi S (1969) The lens capsule. Studies on the carbohydrate units. J Biol Chem 244: 2049-2058.
72. Spiro RG (1969) Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem 244: 602-612.
73. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T (2009) Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol 29: 943-952.
74. Rasmussen M, Jacobsson M, Bjorck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 278: 32313-32316.
75. Wang C, St Leger RJ (2006) A collagenous protective coat enables Metarhizium anisopliae to evade insect immune responses. Proc Natl Acad Sci U S A 103: 6647-6652.
76. Rasmussen M, Eden A, Bjorck L (2000) SclA, a novel collagen-like surface protein of Streptococcus pyogenes. Infect Immun 68: 6370-6377.
77. van Hulten MC, Witteveldt J, Peters S, Kloosterboer N, Tarchini R, et al. (2001) The white spot syndrome virus DNA genome sequence. Virology 286: 7-22.
78. Zhang QY, Xiao F, Xie J, Li ZQ, Gui JF (2004) Complete genome sequence of lymphocystis disease virus isolated from China. J Virol 78: 6982-6994.
79. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, et al. (2010) Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. J Virol 84: 8829-8838.
80. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, et al. (2012) Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). J Biol Chem 287: 3009-3018.
81. Wang Y, Xu Y, Perepelov AV, Qi Y, Knirel YA, et al. (2007) Biochemical characterization of dTDP-D-Qui4N and dTDP-D-Qui4NAc biosynthetic pathways in Shigella dysenteriae type 7 and Escherichia coli O7. J Bacteriol 189: 8626-8635.
82. Dong S, McPherson SA, Wang Y, Li M, Wang P, et al. (2010) Characterization of the enzymes encoded by the anthrose biosynthetic operon of Bacillus anthracis. J Bacteriol 192: 5053-5062.
83. Takeuchi K, Ono H, Yoshida M, Ishii T, Katoh E, et al. (2007) Flagellin glycans from two pathovars of Pseudomonas syringae contain rhamnose in D and L configurations in different ratios and modified 4-amino-4,6-dideoxyglucose. J Bacteriol 189: 6945-6956.
84. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, et al. (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286: 43701-43709.
85. Backlund J, Carlsen S, Hoger T, Holm B, Fugger L, et al. (2002) Predominant selection of T cells specific for the glycosylated collagen type II epitope (263-270) in humanized transgenic mice and in rheumatoid arthritis. Proc Natl Acad Sci U S A 99: 9960-9965.
86. Snir O, Backlund J, Bostrom J, Andersson I, Kihlberg J, et al. (2012) Multifunctional T cell reactivity with native and glycosylated type II collagen in rheumatoid arthritis. Arthritis Rheum 64: 2482-2488.
87. Silverstein AM (2001) Autoimmunity versus horror autotoxicus: the struggle for recognition. Nat Immunol 2: 279-281.
88. Ermann J, Fathman CG (2001) Autoimmune diseases: genes, bugs and failed regulation. Nat Immunol 2: 759-761.

89. Goodnow CC, Sprent J, Fazekas de St Groth B, Vinuesa CG (2005) Cellular and genetic mechanisms of self tolerance and autoimmunity. Nature 435: 590-597.
90. Rioux JD, Abbas AK (2005) Paths to understanding the genetic basis of autoimmune disease. Nature 435: 584-589.
91. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, et al. (2003) Predominant autoantibody production by early human B cell precursors. Science 301: 1374-1377.
92. Nemazee D, Hogquist KA (2003) Antigen receptor selection by editing or downregulation of V(D)J recombination. Curr Opin Immunol 15: 182-189.
93. Zerrahn J, Held W, Raulet DH (1997) The MHC reactivity of the T cell repertoire prior to positive and negative selection. Cell 88: 627-636.
94. Kamradt T, Mitchison NA (2001) Tolerance and autoimmunity. N Engl J Med 344: 655-664.
95. Jenkins MK, Schwartz RH (1987) Antigen presentation by chemically modified splenocytes induces antigen-specific T cell unresponsiveness in vitro and in vivo. J Exp Med 165: 302-319.
96. Falb D, Briner TJ, Sunshine GH, Bourque CR, Luqman M, et al. (1996) Peripheral tolerance in T cell receptor-transgenic mice: evidence for T cell anergy. Eur J Immunol 26: 130-135.
97. Healy JI, Goodnow CC (1998) Positive versus negative signaling by lymphocyte antigen receptors. Annu Rev Immunol 16: 645-670.
98. Benschop RJ, Aviszus K, Zhang X, Manser T, Cambier JC, et al. (2001) Activation and anergy in bone marrow B cells of a novel immunoglobulin transgenic mouse that is both hapten specific and autoreactive. Immunity 14: 33-43.
99. Brunner T, Mogil RJ, LaFace D, Yoo NJ, Mahboubi A, et al. (1995) Cell-autonomous Fas (CD95)/Fas-ligand interaction mediates activation-induced apoptosis in T-cell hybridomas. Nature 373: 441-444.
100. Ravetch JV, Lanier LL (2000) Immune inhibitory receptors. Science 290: 84-89.
101. Sharpe AH, Freeman GJ (2002) The B7-CD28 superfamily. Nat Rev Immunol 2: 116-126.
102. Walker LS, Abbas AK (2002) The enemy within: keeping self-reactive T cells at bay in the periphery. Nat Rev Immunol 2: 11-19.
103. Inobe M, Schwartz RH (2004) CTLA-4 engagement acts as a brake on CD4+ T cell proliferation and cytokine production but is not required for tuning T cell reactivity in adaptive tolerance. J Immunol 173: 7239-7248.
104. Shevach EM (2001) Certified professionals: CD4(+)CD25(+) suppressor T cells. J Exp Med 193: F41-46.
105. Foy TM, Aruffo A, Bajorath J, Buhlmann JE, Noelle RJ (1996) Immune regulation by CD40 and its ligand GP39. Annu Rev Immunol 14: 591-617.
106. Kovanen PE, Leonard WJ (2004) Cytokines and immunodeficiency diseases: critical roles of the gamma(c)-dependent cytokines interleukins 2, 4, 7, 9, 15, and 21, and their signaling pathways. Immunol Rev 202: 67-83.
107. Klein L, Klugmann M, Nave KA, Tuohy VK, Kyewski B (2000) Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. Nat Med 6: 56-61.
108. Pugliese A, Zeller M, Fernandez A, Jr., Zalcberg LJ, Bartlett RJ, et al. (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 15: 293-297.
109. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, et al. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 15: 289-292.
110. Arndt PA, Leger RM, Garratty G (1999) Serology of antibodies to second- and third-generation cephalosporins associated with immune hemolytic anemia and/or positive direct antiglobulin tests. Transfusion 39: 1239-1246.
111. Cloosen S, Arnold J, Thio M, Bos GM, Kyewski B, et al. (2007) Expression of tumor-associated differentiation antigens, MUC1 glycoforms and CEA, in human thymic epithelial cells: implications for self-tolerance and tumor therapy. Cancer Res 67: 3919-3926.

112. Miller SD, Vanderlugt CL, Begolka WS, Pao W, Yauch RL, et al. (1997) Persistent infection with Theiler's virus leads to CNS autoimmunity via epitope spreading. Nat Med 3: 1133-1136.
113. Kamradt T, Soloway PD, Perkins DL, Gefter ML (1991) Pertussis toxin prevents the induction of peripheral T cell anergy and enhances the T cell response to an encephalitogenic peptide of myelin basic protein. J Immunol 147: 3296-3302.
114. Klinman DM, Yi AK, Beaucage SL, Conover J, Krieg AM (1996) CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. Proc Natl Acad Sci U S A 93: 2879-2883.
115. Groux-Degroote S, Krzewinski-Recchi MA, Cazet A, Vincent A, Lehoux S, et al. (2008) IL-6 and IL-8 increase the expression of glycosyltransferases and sulfotransferases involved in the biosynthesis of sialylated and/or sulfated Lewisx epitopes in the human bronchial mucosa. Biochem J 410: 213-223.
116. Lehmann PV, Forsthuber T, Miller A, Sercarz EE (1992) Spreading of T-cell autoimmunity to cryptic determinants of an autoantigen. Nature 358: 155-157.
117. Ferri C, Zignego AL (2000) Relation between infection and autoimmunity in mixed cryoglobulinemia. Curr Opin Rheumatol 12: 53-60.
118. Albert LJ, Inman RD (1999) Molecular mimicry and autoimmunity. N Engl J Med 341: 2068-2074.
119. Ang CW, Jacobs BC, Laman JD (2004) The Guillain-Barre syndrome: a true case of molecular mimicry. Trends Immunol 25: 61-66.
120. Yuki N, Susuki K, Koga M, Nishimoto Y, Odaka M, et al. (2004) Carbohydrate mimicry between human ganglioside GM1 and Campylobacter jejuni lipooligosaccharide causes Guillain-Barre syndrome. Proc Natl Acad Sci U S A 101: 11404-11409.
121. Libbey JE, McCoy LL, Fujinami RS (2007) Molecular mimicry in multiple sclerosis. Int Rev Neurobiol 79: 127-147.
122. Evans CF, Horwitz MS, Hobbs MV, Oldstone MB (1996) Viral infection of transgenic mice expressing a viral protein in oligodendrocytes leads to chronic central nervous system autoimmune disease. J Exp Med 184: 2371-2384.
123. Barzilai O, Ram M, Shoenfeld Y (2007) Viral infection can induce the production of autoantibodies. Curr Opin Rheumatol 19: 636-643.
124. Tagoe CE, Zezon A, Khattri S (2012) Rheumatic manifestations of autoimmune thyroid disease: the other autoimmune disease. J Rheumatol 39: 1125-1129.
125. Kukreja A, Maclaren NK (1999) Autoimmunity and diabetes. J Clin Endocrinol Metab 84: 4371-4378.
126. Kalunian KC, Davis JC, Jr., Merrill JT, Totoritis MC, Wofsy D (2002) Treatment of systemic lupus erythematosus by inhibition of T cell costimulation with anti-CD154: a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 46: 3251-3258.
127. Klareskog L, Catrina AI, Paget S (2009) Rheumatoid arthritis. Lancet 373: 659-672.
128. Zendman AJ, van Venrooij WJ, Pruijn GJ (2006) Use and significance of anti-CCP autoantibodies in rheumatoid arthritis. Rheumatology (Oxford) 45: 20-25.
129. De Rycke L, Peene I, Hoffman IE, Kruithof E, Union A, et al. (2004) Rheumatoid factor and anticitrullinated protein antibodies in rheumatoid arthritis: diagnostic value, associations with radiological progression rate, and extra-articular manifestations. Ann Rheum Dis 63: 1587-1593.
130. Deighton CM, Walker DJ, Griffiths ID, Roberts DF (1989) The contribution of HLA to rheumatoid arthritis. Clin Genet 36: 178-182.
131. Gregersen PK, Silver J, Winchester RJ (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 30: 1205-1213.
132. van der Helm-van Mil AH, Verpoort KN, le Cessie S, Huizinga TW, de Vries RR, et al. (2007) The HLA-DRB1 shared epitope alleles differ in the interaction with smoking and predisposition to antibodies to cyclic citrullinated peptide. Arthritis Rheum 56: 425-432.

133. du Montcel ST, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, et al. (2005) New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. Arthritis Rheum 52: 1063-1068.
134. Van Boxel JA, Paget SA (1975) Predominantly T-cell infiltrate in rheumatoid synovial membranes. N Engl J Med 293: 517-520.
135. Klareskog L, Forsum U, Scheynius A, Kabelitz D, Wigzell H (1982) Evidence in support of a self-perpetuating HLA-DR-dependent delayed-type cell reaction in rheumatoid arthritis. Proc Natl Acad Sci U S A 79: 3632-3636.
136. Lee AT, Li W, Liew A, Bombardier C, Weisman M, et al. (2005) The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. Genes Immun 6: 129-133.
137. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 75: 330-337.
138. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, et al. (2007) Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. J Immunol 179: 4704-4710.
139. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med 357: 1199-1209.
140. Kurreeman FA, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, et al. (2007) A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med 4: e278.
141. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357: 977-986.
142. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39: 1477-1482.
143. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34: 395-402.
144. van der Helm-van Mil AH, Wesoly JZ, Huizinga TW (2005) Understanding the genetic contribution to rheumatoid arthritis. Curr Opin Rheumatol 17: 299-304.
145. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, et al. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 43: 30-37.
146. Silman AJ, Newman J, MacGregor AJ (1996) Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. Arthritis Rheum 39: 732-735.
147. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, et al. (2003) Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis 62: 835-841.
148. Costenbader KH, Feskanich D, Mandl LA, Karlson EW (2006) Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am J Med 119: 503 e501-509.
149. Lofdahl JM, Wahlstrom J, Skold CM (2006) Different inflammatory cell pattern and macrophage phenotype in chronic obstructive pulmonary disease patients, smokers and non-smokers. Clin Exp Immunol 145: 428-437.
150. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, et al. (2008) Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Ann Rheum Dis 67: 1488-1492.
151. Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, et al. (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J Immunol 171: 538-541.

152. Hill JA, Bell DA, Brintnell W, Yue D, Wehrli B, et al. (2008) Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in DR4-IE transgenic mice. J Exp Med 205: 967-979.
153. Klockars M, Koskela RS, Jarvinen E, Kolari PJ, Rossi A (1987) Silica exposure and rheumatoid arthritis: a follow up study of granite workers 1940-81. Br Med J (Clin Res Ed) 294: 997-1000.
154. Stolt P, Kallberg H, Lundberg I, Sjogren B, Klareskog L, et al. (2005) Silica exposure is associated with increased risk of developing rheumatoid arthritis: results from the Swedish EIRA study. Ann Rheum Dis 64: 582-586.
155. Olsson AR, Skogh T, Axelson O, Wingren G (2004) Occupations and exposures in the work environment as determinants for rheumatoid arthritis. Occup Environ Med 61: 233-238.
156. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Berglund G, et al. (2009) Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. Ann Rheum Dis 68: 526-530.
157. Oliver JE, Silman AJ (2006) Risk factors for the development of rheumatoid arthritis. Scand J Rheumatol 35: 169-174.
158. Jonsson IM, Verdrengh M, Brisslert M, Lindblad S, Bokarewa M, et al. (2007) Ethanol prevents development of destructive arthritis. Proc Natl Acad Sci U S A 104: 258-263.
159. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, et al. (2009) Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Ann Rheum Dis 68: 222-227.
160. Albani S, Keystone EC, Nelson JL, Ollier WE, La Cava A, et al. (1995) Positive selection in autoimmunity: abnormal immune responses to a bacterial dnaJ antigenic determinant in patients with early rheumatoid arthritis. Nat Med 1: 448-452.
161. Kim HY, Kim WU, Cho ML, Lee SK, Youn J, et al. (1999) Enhanced T cell proliferative response to type II collagen and synthetic peptide CII (255-274) in patients with rheumatoid arthritis. Arthritis Rheum 42: 2085-2093.
162. Kim WU, Yoo WH, Park W, Kang YM, Kim SI, et al. (2000) IgG antibodies to type II collagen reflect inflammatory activity in patients with rheumatoid arthritis. J Rheumatol 27: 575-581.
163. Kim WU, Cho ML, Jung YO, Min SY, Park SW, et al. (2004) Type II collagen autoimmunity in rheumatoid arthritis. Am J Med Sci 327: 202-211.
164. Edwards JC, Cambridge G, Abrahams VM (1999) Do self-perpetuating B lymphocytes drive human autoimmune disease? Immunology 97: 188-196.
165. Brennan FM, Chantry D, Jackson A, Maini R, Feldmann M (1989) Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. Lancet 2: 244-247.
166. Nishimoto N, Kishimoto T (2006) Interleukin 6: from bench to bedside. Nat Clin Pract Rheumatol 2: 619-626.
167. Burger D, Dayer JM, Palmer G, Gabay C (2006) Is IL-1 a good therapeutic target in the treatment of arthritis? Best Pract Res Clin Rheumatol 20: 879-896.
168. Choy EH, Panayi GS (2001) Cytokine pathways and joint inflammation in rheumatoid arthritis. N Engl J Med 344: 907-916.
169. Ek M, Engblom D, Saha S, Blomqvist A, Jakobsson PJ, et al. (2001) Inflammatory response: pathway across the blood-brain barrier. Nature 410: 430-431.
170. Pettit AR, Ji H, von Stechow D, Muller R, Goldring SR, et al. (2001) TRANCE/RANKL knockout mice are protected from bone erosion in a serum transfer model of arthritis. Am J Pathol 159: 1689-1699.
171. Redlich K, Hayer S, Ricci R, David JP, Tohidast-Akrad M, et al. (2002) Osteoclasts are essential for TNF-alpha-mediated joint destruction. J Clin Invest 110: 1419-1427.
172. Schett G (2007) Erosive arthritis. Arthritis Res Ther 9 Suppl 1: S2.

173. O'Dell JR, Haire CE, Erikson N, Drymalski W, Palmer W, et al. (1996) Treatment of rheumatoid arthritis with methotrexate alone, sulfasalazine and hydroxychloroquine, or a combination of all three medications. N Engl J Med 334: 1287-1291.
174. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, et al. (1997) Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 350: 309-318.
175. Mottonen T, Hannonen P, Leirisalo-Repo M, Nissila M, Kautiainen H, et al. (1999) Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. Lancet 353: 1568-1573.
176. Svensson B, Boonen A, Albertsson K, van der Heijde D, Keller C, et al. (2005) Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial. Arthritis Rheum 52: 3360-3370.
177. Elliott MJ, Maini RN, Feldmann M, Kalden JR, Antoni C, et al. (1994) Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. Lancet 344: 1105-1110.
178. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Davis D, et al. (1998) Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-dose weekly methotrexate in rheumatoid arthritis. Arthritis Rheum 41: 1552-1563.
179. Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, et al. (2000) Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. N Engl J Med 343: 1594-1602.
180. Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, et al. (2003) Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum 48: 35-45.
181. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, et al. (2004) Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 363: 675-681.
182. Bresnihan B, Alvaro-Gracia JM, Cobby M, Doherty M, Domljan Z, et al. (1998) Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. Arthritis Rheum 41: 2196-2204.
183. Smolen JS, Aletaha D, Koeller M, Weisman MH, Emery P (2007) New therapies for treatment of rheumatoid arthritis. Lancet 370: 1861-1874.
184. Kremer JM, Westhovens R, Leon M, Di Giorgio E, Alten R, et al. (2003) Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. N Engl J Med 349: 1907-1915.
185. Edwards JC, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, et al. (2004) Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N Engl J Med 350: 2572-2581.
186. Holmdahl R (2006) Dissection of the genetic complexity of arthritis using animal models. Immunol Lett 103: 86-91.
187. Durie FH, Fava RA, Noelle RJ (1994) Collagen-induced arthritis as a model of rheumatoid arthritis. Clin Immunol Immunopathol 73: 11-18.
188. Holmdahl R, Andersson ME, Goldschmidt TJ, Jansson L, Karlsson M, et al. (1989) Collagen induced arthritis as an experimental model for rheumatoid arthritis. Immunogenetics, pathogenesis and autoimmunity. APMIS 97: 575-584.
189. Brand DD, Latham KA, Rosloniec EF (2007) Collagen-induced arthritis. Nat Protoc 2: 1269-1275.

190. Fugger L, Rothbard JB, Sonderstrup-McDevitt G (1996) Specificity of an HLA-DRB1*0401-restricted T cell response to type II collagen. Eur J Immunol 26: 928-933.
191. Rosloniec EF, Brand DD, Myers LK, Whittington KB, Gumanovskaya M, et al. (1997) An HLA-DR1 transgene confers susceptibility to collagen-induced arthritis elicited with human type II collagen. J Exp Med 185: 1113-1122.
192. Andersson EC, Hansen BE, Jacobsen H, Madsen LS, Andersen CB, et al. (1998) Definition of MHC and T cell receptor contacts in the HLA-DR4restricted immunodominant epitope in type II collagen and characterization of collagen-induced arthritis in HLA-DR4 and human CD4 transgenic mice. Proc Natl Acad Sci U S A 95: 7574-7579.
193. Rosloniec EF, Brand DD, Myers LK, Esaki Y, Whittington KB, et al. (1998) Induction of autoimmune arthritis in HLA-DR4 (DRB1*0401) transgenic mice by immunization with human and bovine type II collagen. J Immunol 160: 2573-2578.
194. Michaelsson E, Andersson M, Engstrom A, Holmdahl R (1992) Identification of an immunodominant type-II collagen peptide recognized by T cells in H-2q mice: self tolerance at the level of determinant selection. Eur J Immunol 22: 1819-1825.
195. Michaelsson E, Malmstrom V, Reis S, Engstrom A, Burkhardt H, et al. (1994) T cell recognition of carbohydrates on type II collagen. J Exp Med 180: 745-749.
196. Fleming SD, Tsokos GC (2006) Complement, natural antibodies, autoantibodies and tissue injury. Autoimmun Rev 5: 89-92.
197. Nandakumar KS, Holmdahl R (2006) Antibody-induced arthritis: disease mechanisms and genes involved at the effector phase of arthritis. Arthritis Res Ther 8: 223.
198. Cho ML, Kang JW, Moon YM, Nam HJ, Jhun JY, et al. (2006) STAT3 and NF-kappaB signal pathway is required for IL-23-mediated IL-17 production in spontaneous arthritis animal model IL-1 receptor antagonist-deficient mice. J Immunol 176: 5652-5661.
199. Hong MH, Williams H, Jin CH, Pike JW (2000) The inhibitory effect of interleukin-10 on mouse osteoclast formation involves novel tyrosine-phosphorylated proteins. J Bone Miner Res 15: 911-918.
200. Takayanagi H, Ogasawara K, Hida S, Chiba T, Murata S, et al. (2000) T-cell-mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN-gamma. Nature 408: 600-605.
201. Stuart JM, Dixon FJ (1983) Serum transfer of collagen-induced arthritis in mice. J Exp Med 158: 378-392.
202. Asquith DL, Miller AM, McInnes IB, Liew FY (2009) Animal models of rheumatoid arthritis. Eur J Immunol 39: 2040-2044.
203. Nandakumar KS, Backlund J, Vestberg M, Holmdahl R (2004) Collagen type II (CII)-specific antibodies induce arthritis in the absence of T or B cells but the arthritis progression is enhanced by CII-reactive T cells. Arthritis Res Ther 6: R544-550.
204. Keystone EC, Schorlemmer HU, Pope C, Allison AC (1977) Zymosan-induced arthritis: a model of chronic proliferative arthritis following activation of the alternative pathway of complement. Arthritis Rheum 20: 1396-1401.
205. van den Broek MF (1989) Streptococcal cell wall-induced polyarthritis in the rat. Mechanisms for chronicity and regulation of susceptibility. APMIS 97: 861-878.
206. Koga T, Kakimoto K, Hirofuji T, Kotani S, Ohkuni H, et al. (1985) Acute joint inflammation in mice after systemic injection of the cell wall, its peptidoglycan, and chemically defined peptidoglycan subunits from various bacteria. Infect Immun 50: 27-34.
207. Schimmer RC, Schrier DJ, Flory CM, Dykens J, Tung DK, et al. (1997) Streptococcal cell wall-induced arthritis. Requirements for neutrophils, P-selectin, intercellular adhesion molecule-1, and macrophage-inflammatory protein-2. J Immunol 159: 4103-4108.
208. Keffer J, Probert L, Cazlaris H, Georgopoulos S, Kaslaris E, et al. (1991) Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis. EMBO J 10: 4025-4031.

209. Kouskoff V, Korganow AS, Duchatelle V, Degott C, Benoist C, et al. (1996) Organ-specific disease provoked by systemic autoimmunity. Cell 87: 811-822.
210. Matsumoto I, Lee DM, Goldbach-Mansky R, Sumida T, Hitchon CA, et al. (2003) Low prevalence of antibodies to glucose-6-phosphate isomerase in patients with rheumatoid arthritis and a spectrum of other chronic autoimmune disorders. Arthritis Rheum 48: 944-954.
211. Koenders MI, Devesa I, Marijnissen RJ, Abdollahi-Roodsaz S, Boots AM, et al. (2008) Interleukin-1 drives pathogenic Th17 cells during spontaneous arthritis in interleukin-1 receptor antagonist-deficient mice. Arthritis Rheum 58: 3461-3470.
212. Geiler T, Kriegsmann J, Keyszer GM, Gay RE, Gay S (1994) A new model for rheumatoid arthritis generated by engraftment of rheumatoid synovial tissue and normal human cartilage into SCID mice. Arthritis Rheum 37: 1664-1671.
213. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1: 2856-2860.
214. Rodriguez-Zaragoza S (1994) Ecology of free-living amoebae. Crit Rev Microbiol 20: 225-241.
215. Wang X, Ahearn DG (1997) Effect of bacteria on survival and growth of Acanthamoeba castellanii. Curr Microbiol 34: 212-215.
216. Weekers PH, Bodelier PL, Wijen JP, Vogels GD (1993) Effects of Grazing by the Free-Living Soil Amoebae Acanthamoeba castellanii, Acanthamoeba polyphaga, and Hartmannella vermiformis on Various Bacteria. Appl Environ Microbiol 59: 2317-2319.
217. Fritsche TR, Horn M, Wagner M, Herwig RP, Schleifer KH, et al. (2000) Phylogenetic diversity among geographically dispersed Chlamydiales endosymbionts recovered from clinical and environmental isolates of Acanthamoeba spp. Appl Environ Microbiol 66: 2613-2619.
218. Rowbotham TJ (1980) Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. J Clin Pathol 33: 1179-1183.
219. Rowbotham TJ (1983) Isolation of Legionella pneumophila from clinical specimens via amoebae, and the interaction of those and other isolates with amoebae. J Clin Pathol 36: 978-986.
220. Rowbotham TJ (1986) Current views on the relationships between amoebae, legionellae and man. Isr J Med Sci 22: 678-689.
221. Svensson L, Jirholt J, Holmdahl R, Jansson L (1998) B cell-deficient mice do not develop type II collagen-induced arthritis (CIA). Clin Exp Immunol 111: 521-526.
222. Jarvinen P, Aho K (1994) Twin studies in rheumatic diseases. Semin Arthritis Rheum 24: 19-28.
223. Svendsen AJ, Holm NV, Kyvik K, Petersen PH, Junker P (2002) Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. BMJ 324: 264-266.
224. Lynn AK, Yannas IV, Bonfield W (2004) Antigenicity and immunogenicity of collagen. J Biomed Mater Res B Appl Biomater 71: 343-354.
225. Yamamoto K (1994) Possible mechanisms of autoantibody production and the connection of viral infections in human autoimmune diseases. Tohoku J Exp Med 173: 75-82.
226. Rees JH, Soudain SE, Gregson NA, Hughes RA (1995) Campylobacter jejuni infection and Guillain-Barre syndrome. N Engl J Med 333: 1374-1379.
227. Cunningham MW, McCormack JM, Fenderson PG, Ho MK, Beachey EH, et al. (1989) Human and murine antibodies cross-reactive with streptococcal M protein and myosin recognize the sequence GLN-LYS-SER-LYS-GLN in M protein. J Immunol 143: 2677-2683.
228. Stollerman GH (1997) Rheumatic fever. Lancet 349: 935-942.
229. McClain MT, Harley JB, James JA (2001) The role of Epstein-Barr virus in systemic lupus erythematosus. Front Biosci 6: E137-147.
230. Poole BD, Scofield RH, Harley JB, James JA (2006) Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. Autoimmunity 39: 63-70.
231. Harley JB, Harley IT, Guthridge JM, James JA (2006) The curiously suspicious: a role for Epstein-Barr virus in lupus. Lupus 15: 768-777.

232. Harley JB, James JA (2006) Epstein-Barr virus infection induces lupus autoimmunity. Bull NYU Hosp Jt Dis 64: 45-50.
233. Swanson-Mungerson M, Longnecker R (2007) Epstein-Barr virus latent membrane protein 2A and autoimmunity. Trends Immunol 28: 213-218.
234. Pender MP (2003) Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases. Trends Immunol 24: 584-588.
235. Costenbader KH, Karlson EW (2006) Epstein-Barr virus and rheumatoid arthritis: is there a link? Arthritis Res Ther 8: 204.
236. Bertoletti A, Gehring AJ (2006) The immune response during hepatitis B virus infection. J Gen Virol 87: 1439-1449.
237. Ichiki Y, He XS, Shimoda S, Ishibashi H, Keeffe EB, et al. (2005) T cell immunity in hepatitis B and hepatitis C virus infection: implications for autoimmunity. Autoimmun Rev 4: 82-95.
238. Bogdanos DP, Mieli-Vergani G, Vergani D (2000) Virus, liver and autoimmunity. Dig Liver Dis 32: 440-446.
239. Gregorio GV, Choudhuri K, Ma Y, Vegnente A, Mieli-Vergani G, et al. (1999) Mimicry between the hepatitis B virus DNA polymerase and the antigenic targets of nuclear and smooth muscle antibodies in chronic hepatitis B virus infection. J Immunol 162: 1802-1810.

Appendix

(Manuscript Shah *et. al.*, 2013 prepared for submission to Journal of Virology)

Exposure to Mimivirus Collagen Promotes Arthritis

Nikunj Shah[a], Andreas J. Hülsmeier[a], Michel Neidhart[b], Steffen Gay[b], Thierry Hennet[a#]

[a]Institute of Physiology and Zurich Center of Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

[b]Center of Experimental Rheumatology, University Hospital Zurich and Zurich Center of Integrative Human Physiology, Gloriastrasse 23, CH-8091 Zurich, Switzerland.

Corresponding author: Thierry Hennet, Institute of Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. Tel: +41-44-635 50 80.

Email: thennet@access.uzh.ch

Running title: Mimivirus collagen and arthritis

Word count abstract: 235

Word count text: 3436

Abstract

Collagens, the most abundant proteins in animals, also occur in some recently described nucleocytoplasmic large DNA viruses such as *Mimiviridae*, which replicate in amoebae. To clarify the impact of viral collagens on the immune response of animals exposed to *Mimiviridae*, we have investigated the localization of collagens in *Acanthamoeba polyphaga* mimivirus particles and the response of mice to immunization with mimivirus particles. Using protein biotinylation, we have first shown that viral collagen encoded by the ORF L71 is present at the surface of mimivirus particles. Exposure to mimivirus collagens elicited the production of anti-collagen antibodies in DBA/1 mice immunized intra-dermally with mimivirus protein extracts. This antibody response also targeted mouse collagen type II and was accompanied by T-cell reactivity to collagen and joint inflammation as observed in collagen-induced arthritis following immunization of mice with bovine collagen type II. The broad distribution of nucleocytoplasmic large DNA viruses in the environment suggests that humans are constantly exposed to such large virus particles. A survey of blood sera from human healthy subjects and from rheumatoid arthritis patients indeed demonstrated that 30% of healthy subject and 36% of rheumatoid arthritis sera recognized the major mimivirus capsid protein L425. Moreover, whereas 6% of healthy subject sera recognized the mimivirus collagen protein L71, 22% of rheumatoid arthritis sera were positive for mimivirus L71. Accordingly, our study shows that environmental exposure to mimivirus represents a risk factor in triggering autoimmunity to collagens.

Introduction

Nucleocytoplasmic large DNA viruses (NCLDVs) represent a growing group of giant viruses found in various types of aquatic environments [16]. NCLDVs include *Poxviridae*, *Asfaviridae*, *Iridoviridae*, *Ascoviridae*, *Phycodnaviridae*, *Mimiviridae,* and *Marseilleviridae* [32]. The *Paramecium bursaria* chlorella virus PBCV-1 was the first large DNA virus characterized at the molecular level and shown to harbor a complex genome of 330 kbp [240]. But the largest NCLDVs described to date belong to *Mimiviridae,* which occur in fresh and saline environments and replicate within amoebae [47]. *Acanthamoeba polyphaga* mimivirus was the first *Mimiviridae* isolated from a cooling water tower and characterized in 2003 [3]. Other members of *Mimiviridae* include megavirus isolated from a marine environment [26], mamavirus [25], and moumouvirus [44]. *Mimiviridae* feature large capsids exceeding 400 nm diameter and harbor large genomes of more than 1 Mbp. The genomes of NCLDVs encode structural proteins and enzymes usually not found in viruses, such as aminoacyl-tRNA synthetases, DNA repair enzymes, potassium ion channel, protein kinases and glycosyltransferases [3,23,241].

Interestingly, *Mimiviridae* also express multiple collagen genes during their infectious life cycle in amoebae. For example, mimivirus expresses seven collagen genes, namely L71, R196, R239, R240, R241, L668, L669, already by 6 h post infection [242]. Even the virophage Sputnik includes two collagen genes among its predicted 21 open reading frames [45]. The functional relevance of these collagens is however presently unknown. First analysis of mimivirus proteins indicated that collagen is hydroxylated like animal collagen [84]. Cryo-electron microscopy and atomic force microscopy studies failed to reveal any collagen-like structures in mimivirus [243,244], although the dense fibers surrounding mimivirus capsids have been suggested to represent cross-linked glycosylated collagen [243].

The ubiquitous distribution of NCLDVs in aquatic environments [54,55] suggests that humans are constantly exposed to such viruses. Mimivirus cannot replicate in animal cells but can be internalized by phagocytosis by mouse and human macrophages [63]. The uptake of mimivirus particles by human macrophages potentially leads to virus antigen presentation and thereby to the generation of antibodies against virus proteins. Considering the structural similarity between animal and *Mimiviridae* collagens, we

made the hypothesis that antibodies generated against *Mimiviridae* collagens may cross-react with animal collagens and thereby contribute to an autoimmune response to collagenous structures in animals previously exposed to *Mimiviridae*. The present study provides evidence supporting this hypothesis, showing that arthritis can be triggered in mice immunized with mimivirus particles and by unravelling increased occurrence of antibodies against mimivirus collagen in rheumatoid arthritis patients.

Materials and Methods

Ethics Statement - All mouse experiments were performed in compliance with the Swiss Animal Protection Ordinance and approved by the local veterinary authority (Kantonales Veterinäramt Zürich, Switzerland). The human sera tested in this study were a part of previously existing collection and the experimental protocol approved by the Kantonale Ethik-Kommission Zürich (KEK).

Mimivirus infection and protein extraction - *Acanthamoeba polyphaga* and mimivirus were provided by Didier Raoult (CNRS UMR6020, Université de la Méditerranée, Marseille). Amoebae were routinely cultured as monolayer in PYG medium at 28°C as previously described [3]. Mimivirus was added to multiplicity MOI 10 to amoebae and newly formed virus was collected from the culture supernatant 2 days post infection. Mimivirus particles were suspended in 0.5 M Tris-HCl, pH 8.5, 0.2% CHAPS, 2 mM TCEP, 6 M guanidine hydrochloride and incubated at 65°C for 10 min. After cooling to room temperature, iodoacetamide was added to a final concentration of 3 mM and further incubated at room temperature for 40 min. After adding DTT to a final concentration of 15 mM, protein extracts were centrifuged at room temperature at 17,000 x g and proteins in the supernatant were precipitated with 12% trichloroacetic acid.

Surface biotinylation of mimivirus proteins - Purified mimivirus particles were suspended in PBS and sulfo-NHS-biotin (Thermo Scientific, Waltham, MA, USA) was added to a final concentration of 1 mg/ml. Virions were rotated for 30 min at room temperature and the reaction was quenched by adding equal volume of 100 mM glycine in PBS. Virions were pelleted and washed twice with 100 mM glycine in PBS and proteins were extracted with guanidine hydrochloride as described above. Extracts were diluted 10-fold in PBS, 0.1 % CHAPS, containing proteinase inhibitors (Calbiochem Proteinase Inhibitor Cocktail III, Merck Millipore, USA) and subjected to avidin cartridge purification (ABSciex, Framingham, Massachusetts, USA). The cartridge was successively washed with 500 µl PBS, 0.1% CHAPS; 1 ml of 650 mM NaCl in 20 mM phosphate buffer pH 7.2, 0.1% CHAPS; 1 ml of PBS, 0.1% CHAPS and 1 ml of 0.1%

CHAPS in H_2O, and biotinylated proteins were eluted with 800 µl 0.4% trifluoroacetic acid, 0.1% CHAPS. Proteins were precipitated with trichloroacetic acid and subjected to SDS-PAGE. Individual protein bands were excised and subjected to in-gel tryptic digestion, as previously described [213,245] followed by LC-MS protein identification. LC-MS data were analyzed using Mascot (Matrix Science, London, UK; version 2.3.02). Mascot was set up to search a Swissprot concatenated target-decoy database (2011.01.11, 1049100 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 PPM. Iodoacetamide derivative of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine and biotinylation of lysine were specified in Mascot as variable modifications. Scaffold (version Scaffold_3.4.9, Proteome Software Inc., Portland, OR, USA) was used to statistically validate MS/MS based peptide and protein identifications.

Collagen induced arthritis model - DBA/1 mice were purchased from Charles River (Germany), bred and maintained in the animal facility of Institute of Physiology, University of Zurich. All experiments were performed in compliance with the Swiss Animal Protection Ordinance and approved by the local veterinary authority (Kantonales Veterinäramt Zürich, Switzerland). Collagen induced arthritis was established as described before [189]. Briefly, 6-8 week old mice were immunized intra-dermally in the tail either with PBS or bovine collagen type II (Chondrex, USA) or mimivirus protein extract emulsified in Complete Freund's Adjuvant (CFA, Chondrex, USA). Each mouse received 50 µl of PBS or 100-120 µg of bovine collagen type II or 120-150 µg of mimivirus proteins emulsified 1:1 in CFA in a total volume of 50 µl. Mice received 30 days later a booster injection of the same amount of antigen emulsified 1:1 in Incomplete Freund's Adjuvant (IFA, Chondrex, USA). Development of arthritis was monitored daily for 75 days post immunization. Severity was scored on a level of 0 (no inflammation) to 4 (most severe inflammation) per limb per mouse, thus allowing a maximum score of 16 per mouse [189].

Anti-collagen type II antibodies - Mouse blood sera were collected by heart puncture. Anti-mouse CII antibody titers were measured in blood sera by ELISA (Chondrex, USA) as per manufacturer's instructions.

Histology – Limbs were skinned and fixed overnight in 10% neutral buffered formalin. Tissues were further decalcified using Immunocal solution (Quartett, Germany) for 4-5 days, dehydrated and paraffin embedded. Sections of 5 µm were mounted on glass slides and stained with H&E.

Recall assay - Axillary, lateral axillary, superficial inguinal and popliteal lymph nodes from mice were collected 8-10 days post booster immunization. Aliquots of 100,000 cells in 100 µl complete RPMI-1640 medium were stimulated with antigens and incubated for 48 h in a CO_2 incubator at 37°C with 5% CO_2. Cells were stimulated in 100 µl either with medium alone as negative control or concanavalin-A (Sigma, Switzerland) at 3 µg/ml as positive control. T-cell proliferation grade denatured mouse collagen type II at 1 mg/ml (Chondrex, USA), T-cell proliferation grade denatured bovine collagen type II at 1 mg/ml (Chondrex, USA) and heat denatured mimivirus collagen L71 at 1.5 mg/ml were used as antigen. After 48 h, 1 µCi of [^{3}H]thymidine (Perkin-Elmer, USA) per well was added, incubated for 16-18 h and cells were harvested on 96-well glass filter (Perkin-Elmer, USA). Radioactivity was counted using a 96-well scintillation beta-counter (Wallac, Perkin-Elmer).

Anti-mimivirus ELISA - Mimivirus proteins were coated in microtiter plates at 0.1 µg per well in 100 µl PBS overnight at 4°C. Plates were washed thrice with PBS-0.05% Tween and blocked with PBS, 0.05% Tween, 1% bovine serum albumin at 37°C for 2 h. Plates were washed, 100 µl of diluted human and rabbit sera added and further incubated at room temperature for 1 h. After three wash steps, 100 µl of 1:5000-diluted biotinylated anti-human or anti-rabbit IgG antibody (BD Biosciences, Switzerland) were added for 2 h. Plates were washed and 100 µl of 1:1000-diluted streptavidin-HRP conjugate (BD Biosciences) added for 1 h in dark. Plates were washed, incubated for 2

min with 50 µl 3, 3' , 5, 5' tetramethylbenzidine (TMB) substrate (BD Biosciences) before stopping the reaction with 25 µl of 2 N H_2SO_4. Colour development was measured at 440 nm.

Immunoprecipitation of mimivirus proteins – Aliquots of 25 µl of human sera were incubated with 30 µl of protein-G sepharose 4 Fast Flow (GE Healthcare, Switzerland) beads along with 80 µl PBS on a rotating shaker for 1 h at 4°C. After centrifugation at 500 x g for 5 min at 4°C, supernatants were discarded and beads incubated with 20 µg of mimivirus protein extract in 80 µl of PBS and further incubated on a rotating shaker for 30 min at 4°C. Beads were washed three times in PBS and antigen-antibody complexes were eluted from the beads by adding 40 µl of 0.1 M glycine, pH 2.7. After neutralization by addition of 20 µl of 1 M Tris-HCl pH 9, eluates were separated by SDS-PAGE. Slices of polyacrylamide gel excluding IgG chains were excised and subjected to in-gel tryptic digest as previously described [213] and peptides identified by tandem mass spectrometry as above.

Cloning, bacterial expression and purification – The mimivirus ORF L71 and L425 were custom synthesized (Genescipt, USA) and subcloned into the expression vector pET16b (Merck Millipore, Switzerland) linearized with *Xho*I and *Hind*III (for L71) or *Xho*I and *BamH*I (for L425). His_6-tagged recombinant proteins were expressed after transformation into *E. coli* BL21 (DE3) cells (Novagen, Switzerland) under induction of 0.2 mM IPTG at 32°C for 1.5 h. Recombinant proteins were purified over Ni-Sepharose 6 Fast Flow (GE Healthcare, Switzerland) gravity flow columns.

Western blotting – Aliquots of 15 µg of recombinant mimivirus L71 and L425 proteins were subjected to SDS-PAGE and transferred to PVDF membrane (Bio-Rad). A His_6-tagged 15 kDa fragment of the human GLT25D2 protein was used as negative control [73]. A 36 kDa His_6-tagged fragment of human collagen type III encompassing 114 [G-X-Y] repeats and lacking N- and C-propeptides was provided by Christoph Rutschmann, Institute of Physiology, University of Zurich. Blots were blocked in 1%

polyvinylpyrrolidone (Sigma, Switzerland) + 5% dry milk solution overnight at 4°C, then washed three times for 5 min with TBS, 0.1% Tween and incubated with human sera diluted 1:4000 for 2 h at room temperature. After washing four times for 5 min, blots were incubated with anti-human IgG-HRP (Promega, Switzerland) at 1:7500 dilution at room temperature for 1 h. Blots were developed with SuperSignal chemiluminescent substrate (Thermo Scientific).

Statistical analysis - One way ANOVA with Dunnett's multiple comparison (GraphPad Prism) was performed to compare experimental groups.

Results

Surface localization of mimivirus collagen

To assess the possible localization of collagens at the surface of mimivirus, we have applied a biotinylation approach on mimivirus particles. Biotinylated mimivirus proteins were captured on streptavidin beads, eluted and identified by tandem mass spectrometry, which revealed 60 surface proteins including the collagen protein L71 (Table S1). The L71 protein has 945 amino acids with 4 collagen domains encompassing 561 amino acids (Fig. 1). The first collagenous domain of L71 includes a stretch with 73% sequence identity to a major human collagen type II T-cell epitope identified in rheumatoid arthritis [85,86]. Other identified surface proteins comprised the capsid protein L425, the putative GMC-type oxidoreductase R135, the thioredoxin domain-containing protein R362 among several uncharacterized proteins (Table S1).

Mimivirus proteins promotes arthritis in mice

Considering the surface expression of collagen L71, we have addressed the potential of mimivirus to induce joint inflammation in DBA/1 mice using the standard protocol for collagen-induced arthritis [189], which closely resembles rheumatoid arthritis in humans. Bovine collagen type II was used as positive control to induce arthritis. Intradermal immunization of bovine collagen type II and mimivirus protein extracts lead to joint inflammation as assessed by visual inspection and histological examination of limb tissues. Mice immunized with mimivirus proteins reached clinical scores of 6 whereas those immunized with bovine collagen type II reached 12 by 75 days (Fig. 2A). Altered cartilage integrity and synovial hyperplasia were evident in joints of mice immunized with bovine collagen type II and to a lesser extent in mice immunized with mimivirus proteins (Fig. 2B).

The breakage of immune tolerance induced by bovine collagen and mimivirus proteins was confirmed by detecting elevated serum titers of anti-mouse collagen type II IgG in both groups of mice immunized with bovine collagen type II and mimivirus proteins (Fig. 3). Furthermore, the cross-reactivity of T-cells was investigated in recall assays [246]. Cells isolated from draining lymph nodes of the immunized mice were found to proliferate in response to *in vitro* presentation of denatured fragments of mouse collagen type II (Fig. 4A), bovine collagen type II (Fig. 4B), and mimivirus collagen

protein L71 (Fig. 4C), thereby confirming the presence of auto-reactive T-cells after immunization with mimivirus proteins. The proliferative response to collagen was strongest for cells isolated from mice immunized with mimivirus proteins. This finding was surprising since mice immunized with bovine collagen type II showed the highest score for limb inflammation and for anti-collagen IgG titers. The strong proliferative response of T-cells from mimivirus proteins-immunized mice may reflect the higher antigenicity of mimivirus collagen considering its peptide sequence divergence from mammalian collagen sequences.

Immunity to mimivirus in humans

To determine whether humans are commonly exposed to mimivirus, we first examined the presence of antibodies against mimivirus in 100 healthy subjects by ELISA using whole mimivirus proteins as antigens. Reactivity to mimivirus proteins was variable; 58 human sera showed significant IgG titers in the 5% range of titers observed in the sera of rabbits previously immunized with mimivirus proteins (Fig. S1). To identify the major mimivirus proteins recognized by human sera, we coupled the IgG fraction of sera from healthy subjects and rheumatoid arthritis patients to protein-G sepharose beads, which were further incubated with preparations of mimivirus proteins. Mimivirus proteins retained on the IgG-protein G beads were identified by mass spectrometric peptide sequencing after trypsin digestion. The major capsid protein L425 was found in all samples followed by the putative GMC type oxidoreductase R135 and core protein L410, which were found in seven from ten samples (Table 1). Interestingly, the most frequent mimivirus proteins recognized by human sera were surface proteins according to our surface biotinylation study (Table S1). Mimivirus collagens did not appear among the proteins recognized by sera. This absence may be related to the abundance of lysine in mimivirus collagens, thereby yielding very short tryptic peptides that remained below the detection range of mass spectrometric peptide sequencing. The recognition of multiple mimivirus proteins by human sera confirmed the exposure of humans to mimivirus.

To further validate the occurrence of antibodies against specific mimivirus proteins in human sera were analyzed by Western blot. The major capsid L425 and collagen L71 proteins were expressed as His_6-tagged recombinant proteins in *E. coli* and purified on Ni^{2+}-sepharose columns. Pools of 100 healthy subject sera and 100 rheumatoid arthritis

sera were probed against the recombinant L425 and L71 mimivirus proteins. We examined the reactivity of sera towards surface collagen L71 since this protein was not detected among the mimivirus proteins captured by immobilized serum IgG in our previous experiment. For the 100 healthy subject and 100 rheumatoid arthritis sera tested, respectively 30 and 36 sera recognized the capsid L425 protein (Fig. 5A, Fig. S2). This results confirmed that exposure to mimivirus is common in the human population. The detection of IgG against the mimivirus capsid protein L425 in 30% of tested sera suggests repeated antigenic challenge probably caused by repeated contact with mimivirus. Reactivity of human sera towards mimivirus collagen was more discriminatory. Whereas only 6 healthy subject sera recognized the mimivirus collagen L71, 22 rheumatoid arthritis sera were positive for the mimivirus collagen L71 (Fig. 5B, Fig. S3). To exclude unspecific cross-reactivity of human sera towards polypeptides containing $[G\text{-}X\text{-}Y]_n$ collagen domains, we have tested the recognition of L71-positive sera for a fragment of human collagen type III encompassing 114 [G-X-Y] repeats and lacking N- and C-propeptides. None of the 28 human sera positive for mimivirus L71 did recognize the 36 kDa $[G\text{-}X\text{-}Y]_{114}$ construct (Fig. 6, Fig. S4), thereby demonstrating the specificity of the antibody response to mimivirus L71 collagen. Accordingly, this work confirmed that the reactivity to mimivirus collagen was 3.5 times more frequent in the pool of rheumatoid arthritis sera in comparison to the limited reactivity of sera from healthy subjects.

Discussion

The present study demonstrated that mice generated auto-reactive anti-collagen antibodies after immunization with mimivirus proteins including viral collagens. A possible relationship between exposure to mimivirus collagen and the development of auto-immunity was corroborated by the occurrence of IgG against mimivirus collagen among rheumatoid arthritis patients. These findings suggested that repeated exposure to mimivirus leads to antibody formation to virus collagen and to a breakage of immune tolerance for endogenous collagens. Giant viruses like mimivirus are ubiquitous in the environment [54,55], thereby supporting the frequent contact of humans to such viruses. Mimivirus is most likely ingested through drinking water and taken up by dendritic cells and macrophages lining the gastrointestinal mucosa. Alternatively, virus particles may enter the airways as aerosol and be taken up by alveolar macrophages. In fact, mimivirus can be phagocytized by human and mouse macrophages although the virus cannot replicate in these cells [63]. Along this line, mimivirus infection has been related to pneumonia in isolated cases, although without evidence for virus particles in disease cases [57,247]. This punctuate pathogenicity however does not preclude a more general effect of mimivirus on priming an auto-immune response.

The sequence similarity between a stretch of mimivirus L71 and human collagen type II (Fig. 1) supports a possible cross-reactivity of antibodies due to antigenic mimicry. A similar case of antigenic mimicry occurs in *Campylobacter jejuni* infection, which causes gastroenteritis, but can lead to Guillain-Barré syndrome when antibodies against *Campylobacter* lipooligosaccharides cross-react with endogenous GM1 gangliosides on nerves cells [120]. Likewise, the detection of antibodies towards mimivirus L425 capsid protein in some *Francisella tularensis* infected patients suggested cross-reactivity of mimivirus antigens with other microorganisms [248]. But surprisingly, no reactivity to mimivirus antigens was found in sera from healthy subjects in the study of Pelletier *et al.* [248].

The main factors involved in the pathogenicity of rheumatoid arthritis could either be genetic or environmental factors. Rheumatoid arthritis is an autoimmune disease with a significant environmental component as supported by twin studies [222,223]. Repeated contact to collagen antigens found in the environment may promote the development of

cross-reactive anti-collagen antibodies and to inflammation in collagen-rich tissues. Antibodies against collagens can either recognize the triple helical conformation, or peptides sequences in the triple helical domain or in telopeptides. The specificity of antibodies against collagen depends on the activation of a humoral response alone, or on a combination of cell-mediated and humoral responses [224]. The reactivity of human sera to mimivirus collagen L71 shown by Western blotting indicates that epitopes based on amino acid sequence and not 3D conformations are being recognized. This notion was supported by finding no reactivity of the L71-positive human sera for the human collagen type III $[G\text{-}X\text{-}Y]_{114}$ polypeptide used as negative control. The lack of recognition for the $[G\text{-}X\text{-}Y]_{114}$ polypeptide also indicated that the reactivity towards L71 was specific to mimivirus exposure and not the result of cross-reactivity to collagen-domain containing proteins such as those found in some Gram-positive bacteria.

Altogether, in view of the structural similarity between mimivirus collagen and human collagen sequences, we propose that giant viruses expressing collagen represent a potential environmental risk factor contributing to the development of rheumatoid arthritis. A systematic survey of mimivirus distribution in the environment will contribute to a better appreciation of the environmental risk associated to such giant viruses in relation to the geographical incidence of rheumatoid arthritis.

Acknowledgements

We are grateful to Lubor Borsig, Daniel Rhyner, Carole Oertli, Jürg Cabalzar, Christoph Rutschmann and Charlotte Burger for technical assistance. We are grateful to the Functional Genomics Center Zurich for its support with mass spectrometric analysis. This work was supported by a Swiss National Science Foundation Grant 310030-129633 to TH.

References

3. Raoult D, Audic S, Robert C, Abergel C, Renesto P, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306: 1344-1350.
16. Wilson WH, Van Etten JL, Allen MJ (2009) The Phycodnaviridae: the story of how tiny giants rule the world. Curr Top Microbiol Immunol 328: 1-42.
23. Wang IN, Li Y, Que Q, Bhattacharya M, Lane LC, et al. (1993) Evidence for virus-encoded glycosylation specificity. Proc Natl Acad Sci U S A 90: 3840-3844.
25. Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, et al. (2011) Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. Genome Biol Evol 3: 737-742.
26. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A 108: 17486-17491.
32. Colson P, Pagnier I, Yoosuf N, Fournous G, La Scola B, et al. (2013) "Marseilleviridae", a new family of giant viruses infecting amoebae. Arch Virol 158: 915-920.
44. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, et al. (2012) Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol Evol 4: 1324-1330.
45. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, et al. (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455: 100-104.
47. Moliner C, Fournier PE, Raoult D (2010) Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev 34: 281-294.
54. La Scola B, Campocasso A, N'Dong R, Fournous G, Barrassi L, et al. (2010) Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology 53: 344-353.
55. Boughalmi M, Saadi H, Pagnier I, Colson P, Fournous G, et al. (2012) High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. Environ Microbiol.
57. La Scola B, Marrie TJ, Auffray JP, Raoult D (2005) Mimivirus in pneumonia patients. Emerg Infect Dis 11: 449-452.
63. Ghigo E, Kartenbeck J, Lien P, Pelkmans L, Capo C, et al. (2008) Ameobal pathogen mimivirus infects macrophages through phagocytosis. PLoS Pathog 4: e1000087.
73. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T (2009) Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol 29: 943-952.
84. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, et al. (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286: 43701-43709.
85. Backlund J, Carlsen S, Hoger T, Holm B, Fugger L, et al. (2002) Predominant selection of T cells specific for the glycosylated collagen type II epitope (263-270) in humanized transgenic mice and in rheumatoid arthritis. Proc Natl Acad Sci U S A 99: 9960-9965.
86. Snir O, Backlund J, Bostrom J, Andersson I, Kihlberg J, et al. (2012) Multifunctional T cell reactivity with native and glycosylated type II collagen in rheumatoid arthritis. Arthritis Rheum 64: 2482-2488.
120. Yuki N, Susuki K, Koga M, Nishimoto Y, Odaka M, et al. (2004) Carbohydrate mimicry between human ganglioside GM1 and Campylobacter jejuni lipooligosaccharide causes Guillain-Barre syndrome. Proc Natl Acad Sci U S A 101: 11404-11409.
189. Brand DD, Latham KA, Rosloniec EF (2007) Collagen-induced arthritis. Nat Protoc 2: 1269-1275.
213. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1: 2856-2860.

222. Jarvinen P, Aho K (1994) Twin studies in rheumatic diseases. Semin Arthritis Rheum 24: 19-28.
223. Svendsen AJ, Holm NV, Kyvik K, Petersen PH, Junker P (2002) Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. BMJ 324: 264-266.
224. Lynn AK, Yannas IV, Bonfield W (2004) Antigenicity and immunogenicity of collagen. J Biomed Mater Res B Appl Biomater 71: 343-354.
240. Van Etten JL, Burbank DE, Kuczmarski D, Meints RH (1983) Virus infection of culturable chlorella-like algae and dlevelopment of a plaque assay. Science 219: 994-996.
241. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV (2010) Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. Biochim Biophys Acta 1800: 152-159.
242. Legendre M, Santini S, Rico A, Abergel C, Claverie JM (2011) Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virol J 8: 99.
243. Xiao C, Chipman PR, Battisti AJ, Bowman VD, Renesto P, et al. (2005) Cryo-electron microscopy of the giant Mimivirus. J Mol Biol 353: 493-496.
244. Kuznetsov YG, Xiao C, Sun S, Raoult D, Rossmann M, et al. (2010) Atomic force microscopy investigation of the giant mimivirus. Virology 404: 127-137.
245. Shin BK, Wang H, Yim AM, Le Naour F, Brichory F, et al. (2003) Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. J Biol Chem 278: 7607-7616.
246. Becher B, Durell BG, Miga AV, Hickey WF, Noelle RJ (2001) The clinical course of experimental autoimmune encephalomyelitis and inflammation is controlled by the expression of CD40 within the central nervous system. J Exp Med 193: 967-974.
247. Vanspauwen MJ, Franssen FM, Raoult D, Wouters EF, Bruggeman CA, et al. (2012) Infections with mimivirus in patients with chronic obstructive pulmonary disease. Respir Med.
248. Pelletier N, Raoult D, La Scola B (2009) Specific recognition of the major capsid protein of Acanthamoeba polyphaga mimivirus by sera of patients infected by Francisella tularensis. FEMS Microbiol Lett 297: 117-123.
249. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74: 5383-5392.
250. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75: 4646-4658.

Table 1. Mimivirus proteins recognized by human serum IgG

ORF	Protein annotation	Healthy subjects					Rheumatoid Arthritis				
L425	Capsid protein	68	601	463	68	387	983	271	258	70	327
R135	Putative GMC-type oxidoreductase		370	394		153	774	255	163		161
L410	Core protein		223	284		179	685	123	133		97
R345	Putative regulator of chromosome condensation	52	123	145	28	58	350	185	77		217
R349	Uncharacterized protein	28	28	30	26	26	35	26		28	32

Scores for viral proteins recognized by 5 healthy subject sera and 5 rheumatoid arthritis sera are listed in columns for each serum tested. Values indicate Mascot scores representing the probability of positive matches for the recognized proteins. Scores above 25 were significant for $p<0.05$.

Supplementary Material

Table S1. Mimivirus surface proteins identified by biotinylation.

ORF	Protein annotation	% coverage
R459	Uncharacterized protein	92%
L724	Uncharacterized protein	69%
L725	Uncharacterized protein	51%
R489	Uncharacterized protein	49%
R714	Uncharacterized protein	42%
L485	Uncharacterized protein	41%
L330	Uncharacterized protein	36%
R305	Uncharacterized protein	36%
R727	Uncharacterized protein	35%
R345	Uncharacterized protein	35%
L53	Uncharacterized protein	33%
R457	Uncharacterized protein	33%
R362	Thioredoxin domain-containing protein	26%
L488	Uncharacterized protein	24%
R346	Uncharacterized protein	24%
L586	Uncharacterized protein	23%
L829	Uncharacterized protein	23%
R623	Uncharacterized protein	22%
L550	Uncharacterized protein	22%
L425	Capsid protein-1	19%
L647	Uncharacterized protein	19%
R306	Uncharacterized protein	19%
L645	Uncharacterized protein	19%

L591	Uncharacterized protein	19%
L719	Uncharacterized protein	18%
R463	Uncharacterized protein	18%
R705	Uncharacterized protein	17%
L585	Uncharacterized protein	17%
L454	Uncharacterized protein	17%
R653	Uncharacterized protein	16%
R610	Uncharacterized protein	14%
L629	Uncharacterized protein	14%
R710	Uncharacterized protein	13%
L399	Uncharacterized protein	12%
R443	Thioredoxin domain-containing protein	12%
R253	Uncharacterized protein	11%
L324	Uncharacterized protein	11%
R307	PP2C-like domain-containing protein	11%
L492	Uncharacterized protein	11%
R596	Probable FAD-linked sulfhydryl oxidase	10%
R135	Putative GMC-type oxido reductase	9.5%
L442	Uncharacterized protein	9.5%
R347	Uncharacterized protein	9.5%
R622	Putative tyrosine-protein phosphatase	9%
L609	Uncharacterized protein	8.6%
L612	Uncharacterized protein	7.5%
L309	Uncharacterized protein	6.2%
R252	Uncharacterized protein	6.1%
R526	Putative alpha/beta hydrolase	6.1%

R692	Uncharacterized protein	4.8%
L236	Uncharacterized protein	4.7%
L448	Uncharacterized protein	4.5%
L264	Uncharacterized WD repeat-containing protein	4.5%
R588	Uncharacterized protein	3.3%
L605	Structural PPIase-like protein	3%
R553	Uncharacterized protein	2.7%
L71	Collagen-like protein 1	2.5%
L357	Uncharacterized protein	2.4%
R643	Uncharacterized protein	2.2%
L397	Uncharacterized protein	2.1%

Mimivirus surface proteins identified by mass spectrometric peptide sequencing after biotinylation and analyzed by Mascot software. Results were validated by Scaffold (version Scaffold_3.4.9, Proteome Software Inc., Portland, OR) and peptide identifications were accepted if they established >80% probability as specified by the Peptide Prophet algorithm [249]. Protein identifications were accepted if they could be established at >99% probability and contained at least 2 identified peptides as specified by Protein Prophet algorithm [250].

Figure 1

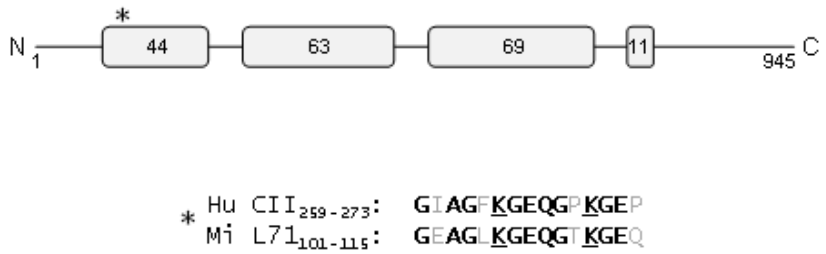


Figure 1. Domain organization of mimivirus L71 protein. The four collagen domains of L71 are shown as grey boxes with the number of G-X-Y repeats given inside. The asterisk shows the position of the sequence motif similar to the epitope human collagen type II recognized as immunodominant in rheumatoid arthritis [85]. The sequence of this human collagen type II (Hu CII) epitope encompassing amino acids 259-273 is shown aligned with the corresponding sequence of mimivirus L71 (Mi L71) encompassing amino acids 101-115.

Figure 2

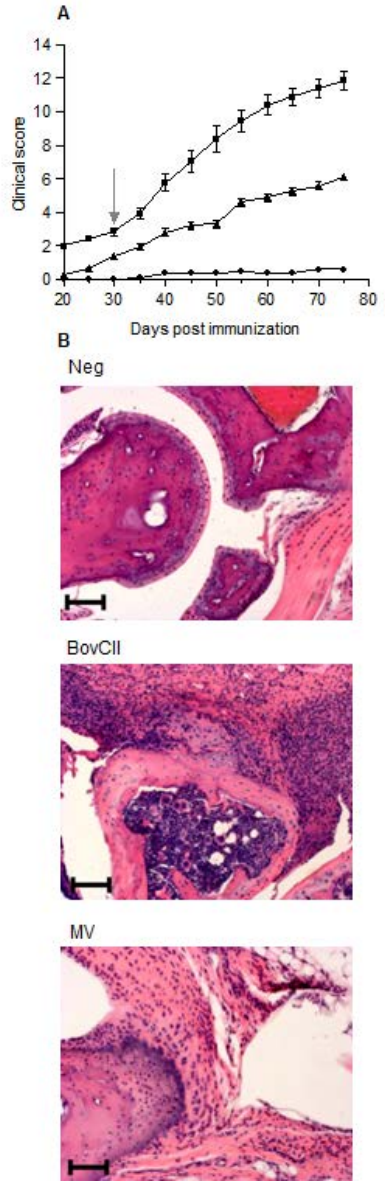


Figure 2. Joint inflammation in DBA/1 mice immunized with mimivirus proteins. Mice were immunized with either PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). (**A**) The clinical severity of arthritic limbs in the groups of Neg (circles), BovCII (squares) and MV (triangles) immunized mice are shown as mean ± SEM. The arrow shows the time point of booster immunization. Data represent three independent experiments including 15-21 mice per group. (**B**) Representative H&E stained sections of hind limbs by day 75 after immunization showing cartilage damage and synovial hyperplasia in BovCII and MV immunized mice. No sign of pathology were visible in negative control mice (Neg), scale bar 100 μm.

Figure 3

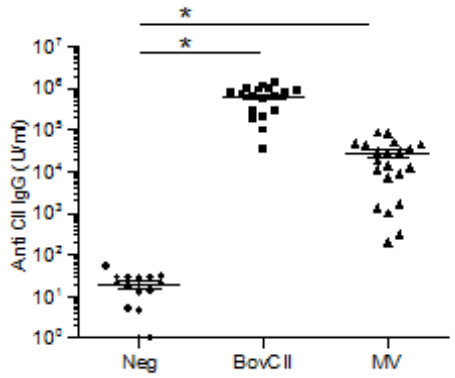


Figure 3. **Anti-collagen type II IgG titers in DBA/1 mice immunized with mimivirus proteins.** Levels of serum IgG measured by ELISA against endogenous mouse collagen type II (CII) in mice immunized with PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV). Data represent three independent experiments including 15-21 mice per group, horizontal bars show mean ±SEM, *p<0.01.

Figure 4

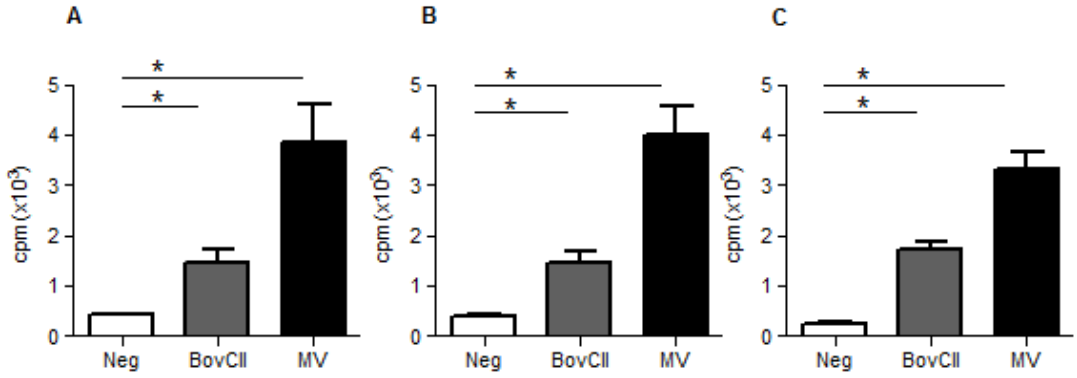


Figure 4. **Auto-reactive T-cell response in DBA/1 mice immunized with mimivirus proteins.** (**A**) Recall responses in cells isolated from draining lymph nodes from mice immunized with PBS (Neg), bovine collagen type II (BovCII), or mimivirus proteins (MV) after stimulation with denatured mouse collagen type II. (**B**) Recall responses after stimulation with denatured bovine collagen type II. (**C**) Recall responses after stimulation with denatured fragmented recombinant mimivirus protein L71. Data represent mean ± SEM from groups of 3 mice, *p<0.01.

Figure 5

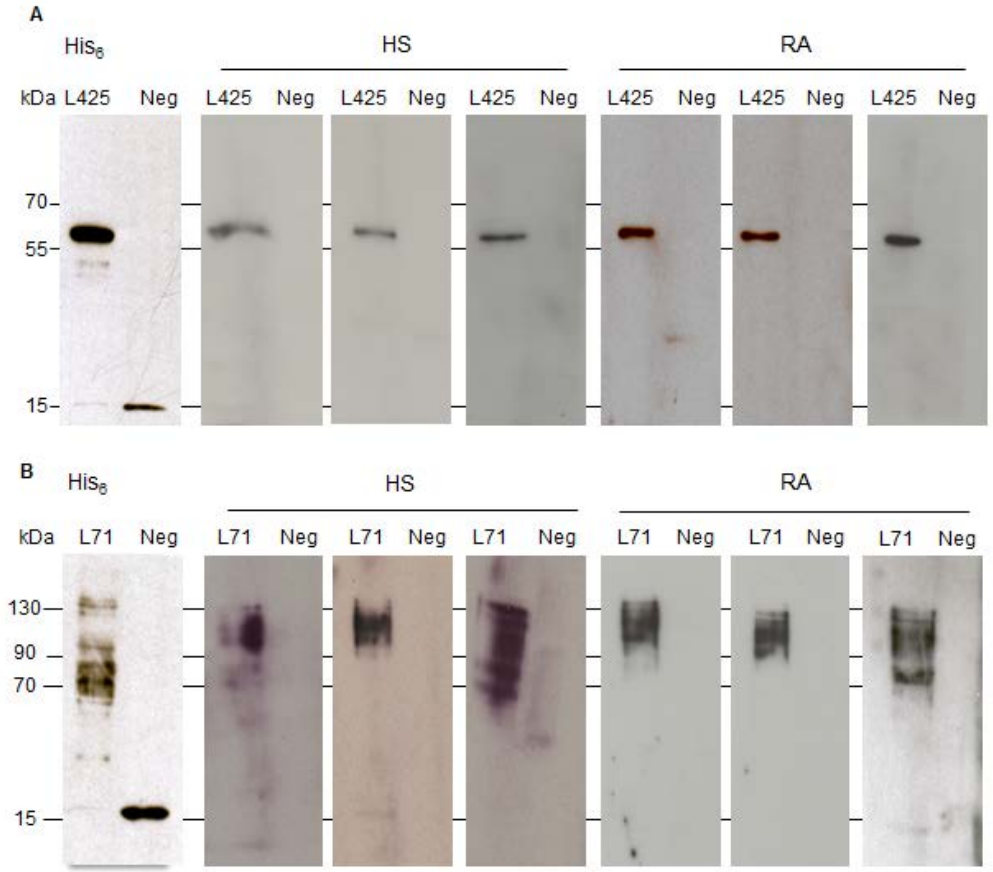


Figure 5. Recognition of mimivirus proteins by human sera. (**A**) Representative Western blots of sera from healthy subjects (HS) and rheumatoid arthritis (RA) patients recognizing mimivirus capsid protein L425. (**B**) Representative Western blots of sera from healthy subjects (HS) and rheumatoid arthritis (RA) patients recognizing mimivirus collagen L71. Sera were diluted 1:4000. Positions of recombinant L425 and L71 proteins in the blots are shown at the left of each panel using an anti-His_6 antibody (His_6). A 15 kDa fragment of His_6-tagged human GLT25D2 protein was used as negative control (Neg).

Figure 6

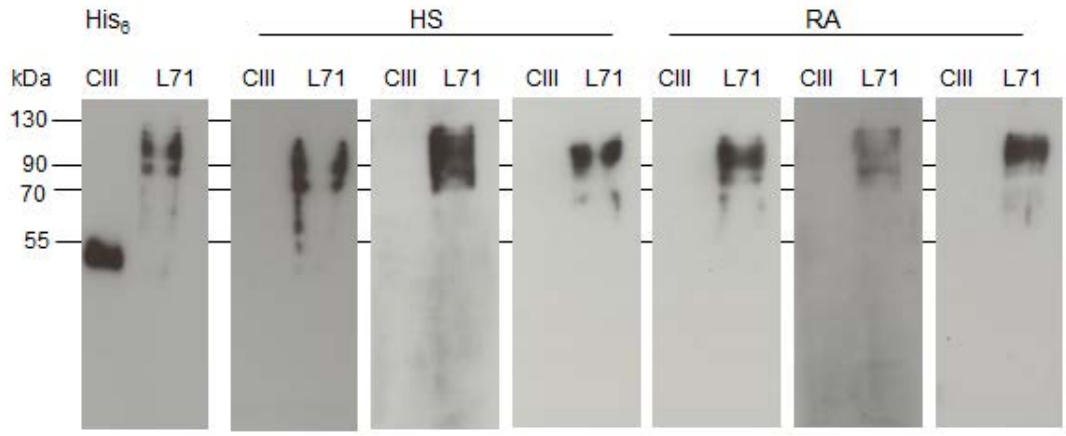


Figure 6. Specific recognition of mimivirus collagen L71 by human sera. Representative Western blots of L71-positive sera from healthy subjects (HS) and rheumatoid arthritis (RA) patients recognizing mimivirus collagen L71 but not a fragment of human collagen type III containing 114 [G-X-Y] repeats (CIII). Sera were diluted 1:4000. Positions of recombinant L71 and CIII proteins in the blots are shown at the left of the panel using an anti-His_6 antibody (His_6).

Supplementary Figure S1

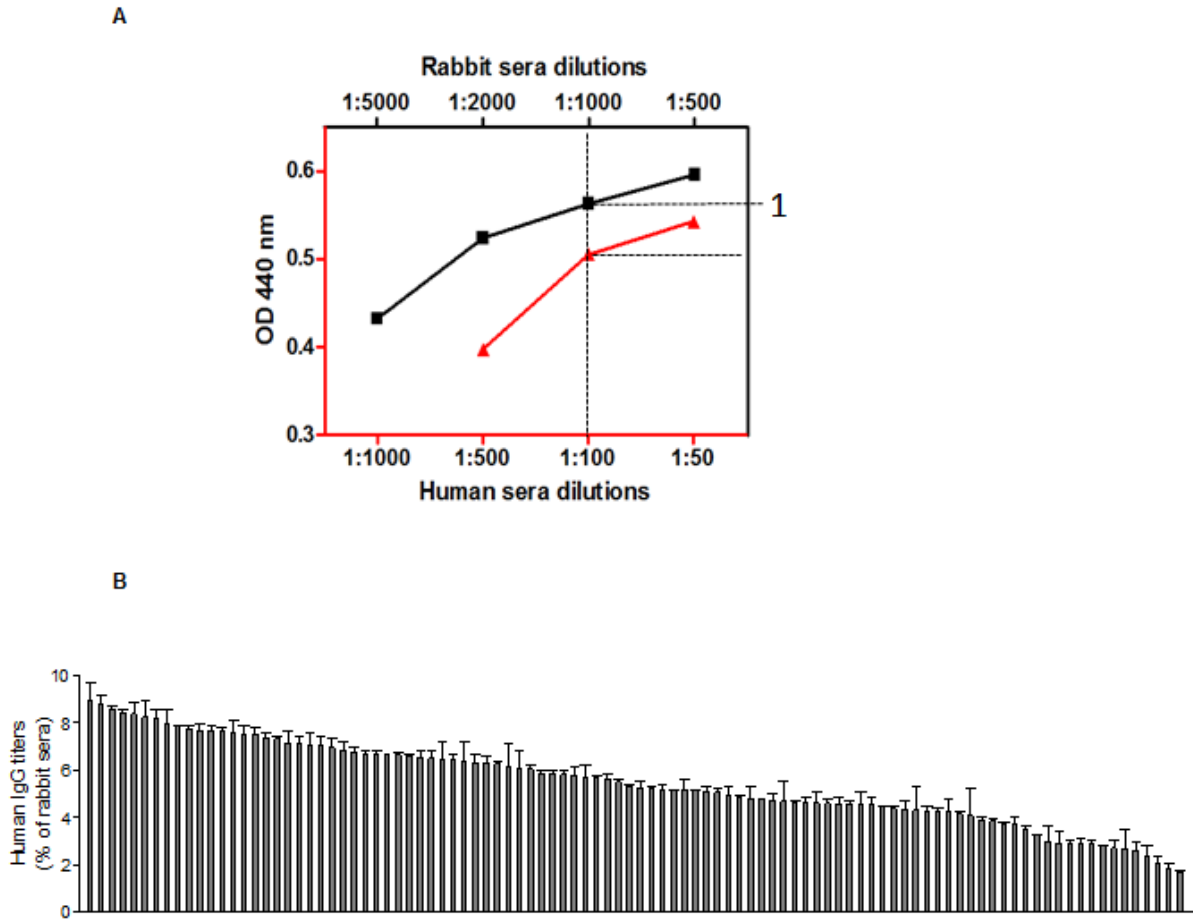


Figure S1. Reactivity of human sera towards mimivirus proteins. (**A**) Representative plot showing the response of human sera at dilution 1:100 (red line) compared to the response of rabbit sera at dilution 1:1000 (black line). (**B**) Anti-mimivirus IgG titers in sera of 100 healthy subjects were measured by ELISA and expressed as a ratio to IgG titers measured in rabbits previously immunized with mimivirus particles. Error bars show SEM from two independent experiments, performed in duplicates.

Supplementary Figure S2

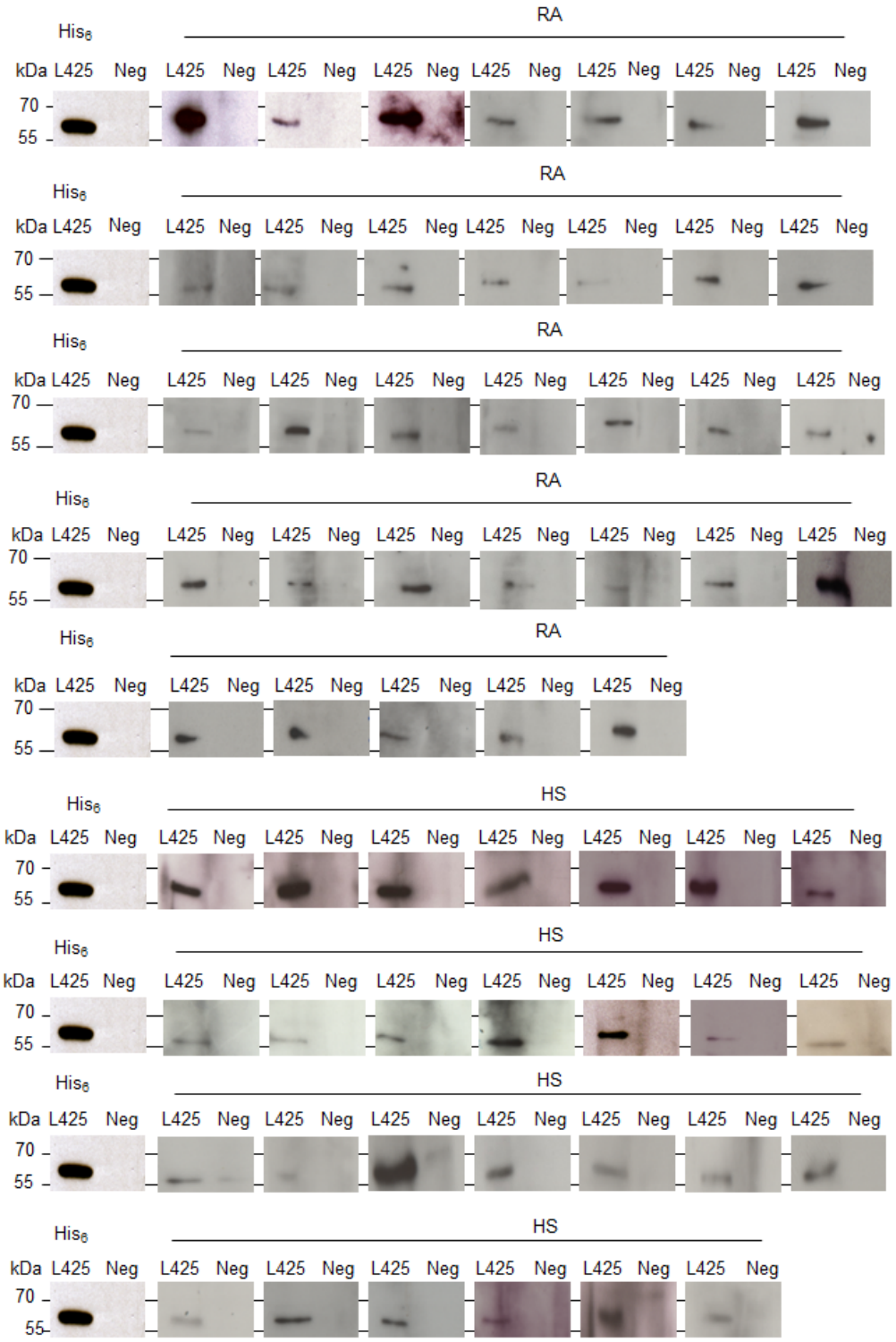


Figure S2. Recognition of mimivirus capsid L425 protein by human sera. Pools of 100 healthy subjects (HS) and 100 rheumatoid arthritis (RA) patient's sera were tested for reactivity with mimivirus capsid protein L425 by Western blotting. 30 HS and 36 RA

sera recognized L425. Representative 3 blots each of HS and RA sera recognizing L425 are shown in Fig. 5A. The remaining Western blots of sera from 27 HS and 33 RA sera recognizing mimivirus capsid protein L425 are shown here. The 70 HS and 64 RA sera which did not show reactivity to L425 are not shown. Sera were diluted 1:4000. Positions of recombinant L425 protein in the blots are shown at the left of each panel using an anti-His_6 antibody (His_6).

Supplementary Figure S3

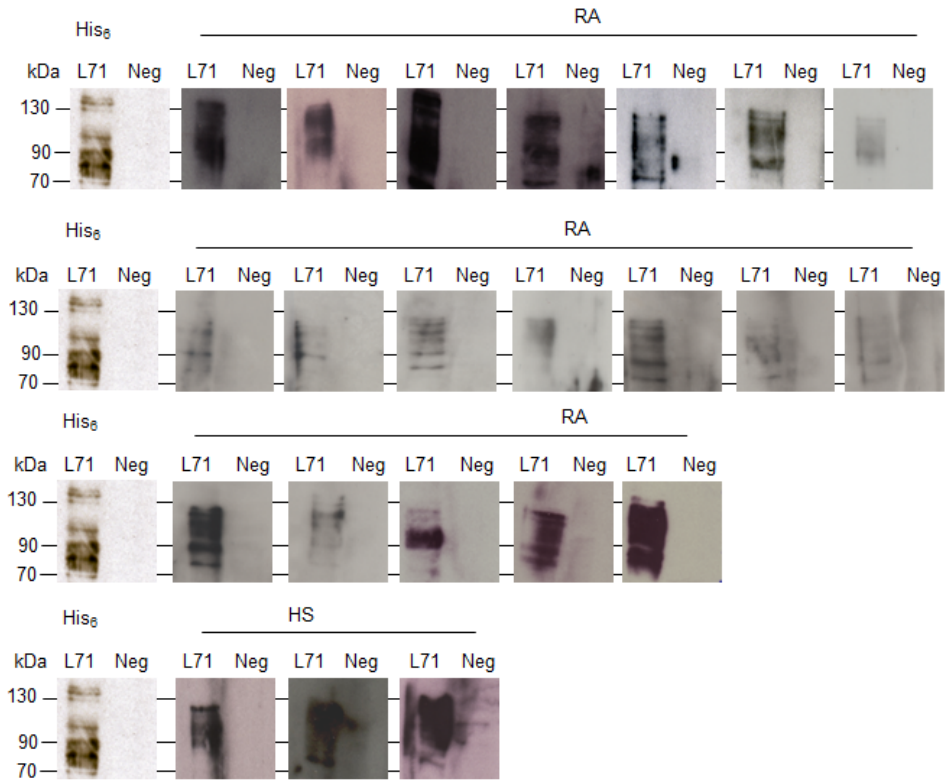


Figure S3. Recognition of mimivirus collagen L71 by human sera. Pools of 100 healthy subjects (HS) and 100 rheumatoid arthritis (RA) patient's sera were tested for reactivity with mimivirus collagen L71 by Western blotting. 6 HS and 22 RA sera recognized L71. Representative 3 blots each of HS and RA sera recognizing L71 are shown in Fig. 5B. The remaining western blots of sera from 3 HS and 19 RA sera recognizing mimivirus collagen L71 are shown here. The 94 HS and 78 RA sera which did not show reactivity to L71 are not shown. Sera were diluted 1:4000. Positions of recombinant L71 protein in the blots are shown at the left of each panel using an anti-His_6 antibody (His_6).

Supplementary Figure S4

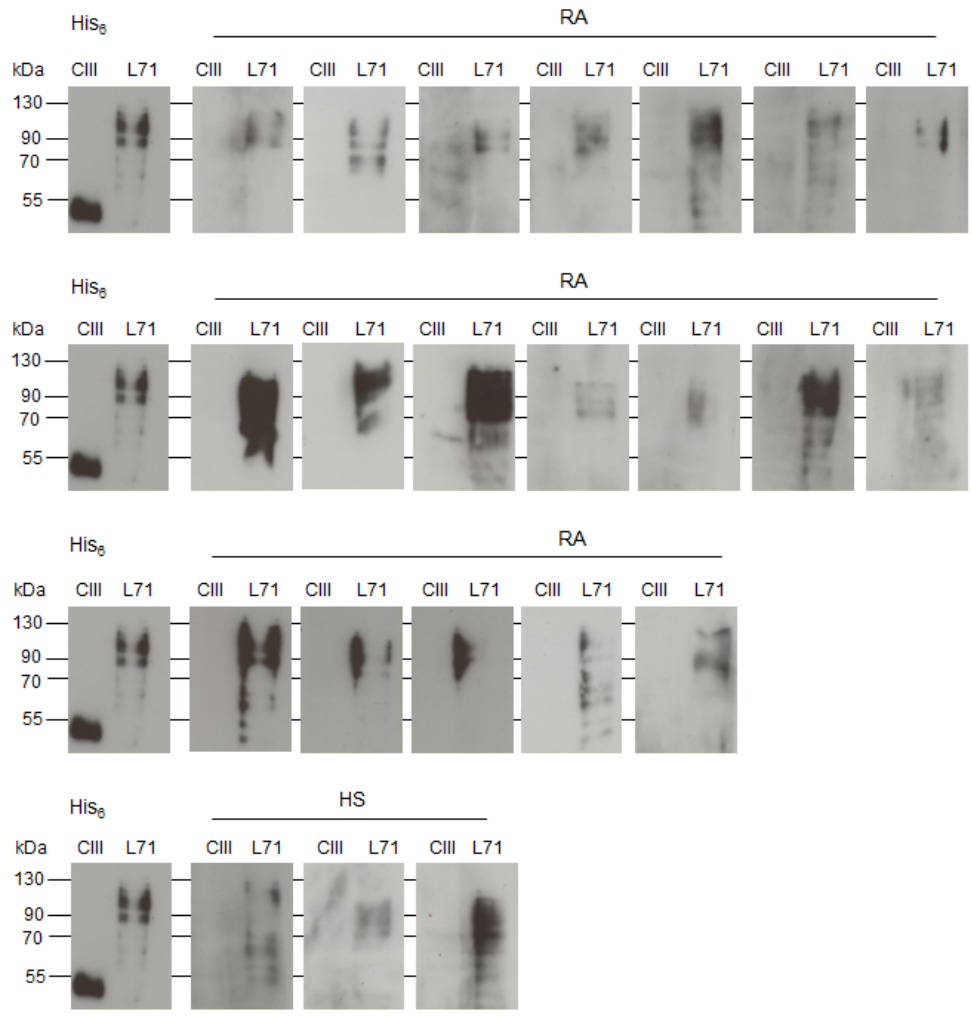


Figure S4. Specific recognition of mimivirus collagen L71 by human sera. Western blots of L71-positive sera from healthy subjects (HS) and rheumatoid arthritis (RA) patients recognizing mimivirus collagen L71 but not a fragment of human collagen type III containing 114 [G-X-Y] repeats (CIII). Three representative blots from each HS and RA sera are shown in Fig. 6. The remaining blots of sera from 3 HS and 19 RA sera recognizing mimivirus collagen L71 but not CIII are shown here. Sera were diluted 1:4000. Positions of recombinant L71 and CIII proteins in the blots are shown at the left of the panel using an anti-His_6 antibody (His_6).

Acknowledgments

At the onset I would sincerely like to thank my supervisor Prof. Dr. Thierry Hennet for accepting me as a grad student in his lab and allowing me to perform this project, it was indeed a pleasure to work in your lab. I would also like to thank him for his constant support, mentoring, suggestions and comments all throughout my work. I hope I was a little bit successive in making you differentiate between real Bollywood movies and other movies from my country ☺

Many thanks to my committee members PD Dr. Lubor Borsig, PD Dr. Patricie Burda and Prof. Dr. Matthias Baumgartner for their valued suggestions and ideas during all the committee meetings. Special thanks to Lubor for helping us during our animal experiments.

I would take this opportunity to thank my master thesis supervisors Dr. Alan McNally and Dr. Georgina Manning from Nottingham Trent University, UK for their very first push and encouragement to motivate me to pursue my doctorate studies.

Special thanks to Carole Oertli for performing the anonymous scoring of our animals, Jürg Cabalzar for always helping analyse mass spectrometric data, Charlotte Burger (Institute of Anatomy) for helping in histology, Steffen Gay and Michel Neidhart (Centre of Experimental Rheumatology) for providing us with human sera samples for our project, Petra Seebeck and Marija Mihailova (Zurich Integrative Rodent Physiology) for their immense help during the animal imaging experiments, Matthew Adjei and Daniel Pochetti for taking care of my animals in the mouse facility, Michael Welti for helping me write the German summary for this thesis, Sacha Schneeberger for help during the virus isolation experiments and Stephan Baumann for his constant help for technology related issues which I faced a lot.

Life would have been very difficult if I didn't have the scientific, emotional and super-friendly support of my dear fellow colleagues on the L-floor. Dr. G Weiss (our dear baking expert), Michi (our 'artial martist' and my fellow Asian counterpart), Jürgiiiiii (my 'dear' friend), Büüme (for being so 'awesome'), Eddie (my other fellow Asian counterpart), Anna (my LSC expert), Sacha (my personal trainer), Christoph (for answering my constant questions and loving my country so much), Dr. Kelvin (for his

constant un-understandable jokes), Jesus (for providing us with the world's best hugs and that too every day), mamma Giovi (for taking care of us like a mommy), Dr. Andreas, Dr. Andrea (my dear Swiss cheese fondue and also my ELISA teacher) and Dr. Charlotte (for her concerts and my very first western blot teacher). A big thank you to everyone from the Borsig group as well- Dr. Katya, Dr. Alexa, Irina, Carole, Marco and Darya for the constant happy working atmosphere. Thank you all for bearing with my annoyingness and not making me feel homesick for most of my time here in Zurich, it was indeed a pleasure seeing all of you every day for the past few years. I am glad to have made some real great 'friends for life' during these years.

Many sincere thanks to the friends I have made outside the working environment with whom I have enjoyed many weekend trips, parties and much more. Thanks to Carmen, Patrick, Nigel, Sarah, Nina, Joao, Shreya, Shilpi, Samyuktha and Amit.

Last but surely not the least, many many thanks to Lord Almighty and my family- mom, dad, brother, bhabhi- this wouldn't have been possible without your emotional and moral support. I am sincerely grateful to you. Thank you ☺